Journal of Econometrics 177 (2013) 250–264

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Predictive regression under various degrees of persistence and robust long-horizon regression

Peter C.B. Phillips [a,b,c,d,*], Ji Hyung Lee [a]

[a] Yale University, United States
[b] University of Auckland, New Zealand
[c] Singapore Management University, Singapore
[d] University of Southampton, United Kingdom

ARTICLE INFO

Article history:
Available online 17 April 2013

JEL classification:
C22

Keywords:
Asymptotic theory
Balanced regression
Endogeneity
Instrumentation
IVX methods
Local power
Mild integration
Mildly explosive
Predictive regression
Robustness

ABSTRACT

The paper proposes a novel inference procedure for long-horizon predictive regression with persistent regressors, allowing the autoregressive roots to lie in a wide vicinity of unity. The invalidity of conventional tests when regressors are persistent has led to a large literature dealing with inference in predictive regressions with local to unity regressors. Magdalinos and Phillips (2009b) recently developed a new framework of extended IV procedures (IVX) that enables robust chi-square testing for a wider class of persistent regressors. We extend this robust procedure to an even wider parameter space in the vicinity of unity and apply the methods to long-horizon predictive regression. Existing methods in this model, which rely on simulated critical values by inverting tests under local to unity conditions, cannot be easily extended beyond the scalar regressor case or to wider autoregressive parametrizations. In contrast, the methods developed here lead to standard chi-square tests, allow for multivariate regressors, and include predictive processes whose roots may lie in a wide vicinity of unity. As such they have many potential applications in predictive regression. In addition to asymptotics under the null hypothesis of no predictability, the paper investigates validity under the alternative, showing how balance in the regression may be achieved through the use of localizing coefficients and developing local asymptotic power properties under such alternatives. These results help to explain some of the empirical difficulties that have been encountered in establishing predictability of stock returns.

1. Introduction

Predictive regression typically encounters the problem of predicting some noisy stationary variable using a highly persistent regressor. A leading practical example is stock return predictability in finance and the empirical puzzles associated with these regressions that have emerged in the financial literature. The traditional form of the efficient market hypothesis supports the idea of martingale behavior in stock prices and stock return unpredictability. But empirical evidence on the predictability of returns shows mixed results on the explanatory power of various economic fundamentals such as the dividend–price ratio, leading to what has become known as the stock return predictability puzzle. Some researchers have even characterized stock return predictability as a new stylized fact in finance.

Forecasting stock returns has been a longstanding interest of Hashem Pesaran. His well cited paper with Alan Timmerman (1995) is an early contribution in the field that highlighted the need to robustify econometric procedures to the time varying predictive power of economic factors on stock returns. The present paper explores a related theme and investigates robust predictive regressions in the presence of multivariate nonstationary regressors developing results that have direct application to stock return forecasting.

The empirical model employed in stock return predictive regressions commonly involves a linear regression of returns on economic fundamentals. The regressors typically manifest a high, but imprecisely determined, degree of persistence. This uncertainty in the degree of regressor persistence is usually modeled in terms of an autoregressive coefficient with an unknown local to unity parameter (the localizing coefficient) that measures the (sample size normalized) departure of this autoregressive coefficient from unity. The localizing coefficient is not consistently estimable and this characteristic leads to nonstandard and nonpivotal inference problems.

Other commonly occurring predictive regressions include forward premium regressions in international finance and consumption growth regressions in macroeconomics. These models share the same problem of nonstandard and nonpivotal inference that

* Corresponding author at: Yale University, United States.
E-mail addresses: peter.phillips@yale.edu (P.C.B. Phillips), jihyung.lee@yale.edu (J.H. Lee).

0304-4076/$ – see front matter
http://dx.doi.org/10.1016/j.jeconom.2013.04.011

originates in the unknown degree of persistence in the predictive regressors. These common difficulties in predictive regressions have kindled a widespread search for robust inference methods. There have been extensive efforts and some procedures such as Bonferroni type methods have received much attention. At present, Bonferroni type procedures represent the state-of-the-art in this literature but they do have some undesirable and limiting properties. In particular, simulations are required to compute critical values to perform inference and confidence interval construction since the limit distribution employed in the calculations is nonstandard. A second limitation is that the method is very difficult to extend beyond the scalar regressor case. In practical work, there are often a selection of variables representing various economic fundamentals which need to be investigated in applied work on predictive regression and it is delimiting for empirical procedures to be restricted to a single regressor.

Magdalinos and Phillips (2009b, MP henceforth) recently developed a novel extended IV procedure (called IVX regression) and established some attractive asymptotic features of this method that apply in quite general cointegrating regression models. In particular, the IVX method has some useful and somewhat surprising features such as standard chi square testing without any precise knowledge about the degree of persistence in the regressors, straightforward extension of the methods to multivariate models, and great generality in terms of the permissible persistence (or vicinity of unity) space. This method has very recently been applied to a predictive regression model and was shown to inherit all these advantages in this context (Kostakis et al., 2010).

In empirical research short horizon predictive regressions have shown generally inconclusive findings. In response, researchers have been studying prediction over longer horizons for such variables as stock returns and the forward premium. Long-horizon predictive regressions share the same problems as their short horizon counterparts (viz., nonstandard and nonpivotal inference with persistent regressors) and similar solutions such as Bonferroni techniques have been applied with the same limitations noted above. It is therefore natural to explore whether IVX methodology has potentially beneficial applications in long-horizon regressions.

That question forms the focus of the present paper. We study long-horizon predictive regressions and propose a novel inference procedure which is based on an extended version of MP and Kostakis et al. (2010). The long-horizon version of IVX is analyzed, and shown to be applicable to even much wider parameter region near unity: from boundary of stationary/unit root side (mildly integrated regressors) to mildly explosive regressors. This is the most extensive treatment of the parameter region near unity not only within this setting of predictive regression but also in more general time series regressions such as cointegrating regressions. All the attractive features of IVX regression such as standard asymptotic chi square inference with possible multivariate regressor and regressand are shown to apply.

A further contribution of the paper is to investigate validity of the predictive model specification under the alternative. In particular, we show how balance in the regression may be achieved through the use of localizing coefficients and that non trivial local asymptotic power applies under such alternatives. These results partly explain some of the practical difficulty that has been encountered in establishing empirical evidence of predictability. In effect, the departures from the null that deliver predictability are necessarily small in order to preserve the observed character of the dependent variable, thereby making detection difficult. Nonetheless, as we show here, long horizon IVX regression provides a simple and effective machinery for testing predictability that has non-trivial asymptotic power against local alternatives.

The paper is organized as follows. Section 2 overviews existing results on predictive regression literature and the various limitations of the methods currently used in empirical research. Section 3 develops a limit theory for the extended IVX approach in long horizon predictive regression. Section 4 concludes and proofs of the main results are given in the Appendix. The Appendix also contains a discussion of balancing predictive regression and an analysis of local asymptotic power.

A supplement (Phillips and Lee, 2012a) is available online and provides supporting lemmas and further technical arguments that are used in the paper.

2. Predictive regressions: literature review and motivation

This Section reviews key results in the predictive regression literature and identifies the source of the difficulties encountered by existing methods. The basic linear predictive model can be characterized as:

$$y_t = \beta x_{t-1} + u_{0t}, \tag{2.1}$$

$$x_t = \rho x_{t-1} + u_{xt}. \tag{2.2}$$

We impose a simple but widely used structure of martingale difference sequence (mds) innovations for $u_t := [u_{0t}, u_{xt}]$ with conditional variance

$$\mathbb{E}_{\mathcal{F}_{t-1}}\left[u_t u_t'\right] = \begin{bmatrix} \Sigma_{00} & \Sigma_{0x} \\ \Sigma_{x0} & \Sigma_{xx} \end{bmatrix},$$

where $\mathcal{F}_t$ is the natural filtration. This framework allows only for contemporaneous correlation between the components of the model. More general dependence structures will be permitted later but for the purpose of this overview we retain the simple mds structure. Full details of the conditions and notation used in the paper are provided in Appendix A.1.

2.1. Existing problems

2.1.1. Finite sample bias with stationary regressors

The centered OLS coefficient estimator in (2.1) has the form:

$$\hat{\beta} - \beta = \frac{\sum_{t=1}^n x_{t-1}u_{0t}}{\sum_{t=1}^n (x_{t-1})^2} = \frac{\sum_{t=1}^n x_{t-1}u_{0.xt}}{\sum_{t=1}^n (x_{t-1})^2} + \left(\frac{\Sigma_{0x}}{\Sigma_{xx}}\right)\frac{\sum_{t=1}^n x_{t-1}u_{xt}}{\sum_{t=1}^n (x_{t-1})^2}$$

$$= \frac{\sum_{t=1}^n x_{t-1}u_{0.xt}}{\sum_{t=1}^n (x_{t-1})^2} + \left(\frac{\Sigma_{0x}}{\Sigma_{xx}}\right)\left(\hat{\rho} - \rho\right), \tag{2.3}$$

where $u_{0.xt} = u_{0t} - \frac{\Sigma_{0x}}{\Sigma_{xx}}u_{xt}$ and $\hat{\rho} = \left(\sum_{t=1}^n x_{t-1}^2\right)^{-1}\sum_{t=1}^n x_{t-1}x_t$. Under normality and with a stationary regressor ($|\rho| < 1$), Stambaugh (1999) gave a bias expansion for $E\left(\hat{\beta} - \beta\right)$ using the well known bias expansion for a fitted AR(1) (Kendall, 1954), which in the case of a fitted intercept has the form

$$E\left(\hat{\beta} - \beta\right) = -\left(\frac{\Sigma_{0x}}{\Sigma_{xx}}\right)\left(\frac{1+3\rho}{n}\right) + O\left(\frac{1}{n^2}\right), \tag{2.4}$$

which has come to be known as the "Stambaugh bias".[1] Accordingly, the first order bias adjusted estimator has been used by

[1] As indicated, the bias formula given in (2.4) is for a stationary AR(1) process with a fitted intercept. Unlike the stationary case, fitting an intercept affects asymptotics in both the nonstationary and explosive regressor cases. For the subsequent development, which focuses on persistent and explosive regressors, it is convenient to keep to the no-intercept case in the generating mechanism for x_t. On the other hand, introducing an intercept in the predictive regression (2.1), so that $y_t = \mu_y + \beta x_{t-1} + u_{0t}$, is easily handled even with nonstationary regressors — see Kostakis et al. (2010) — and this is the primary case of interest in practice.

subsequent researchers, e.g., Kothari and Shanken (1997),

$$\hat{\beta}_{adj} = \hat{\beta} + \frac{\hat{\Sigma}_{0x}}{\widehat{\Sigma}_{xx}}\left(\frac{1+3\hat{\rho}}{n}\right), \tag{2.5}$$

and Amihud and Hurvich (2004) refined this estimator by using a second order bias correction. Within this framework, the problem is only considered as a "finite sample problem", and therefore disappears asymptotically. Additionally, the bias formula (2.4) and associated correction (2.5) is only valid in the stationary case.

2.1.2. Nonstandard limit theory and uncorrectable bias with persistent regressors

There is wide consensus that most economic fundamentals used as regressors in predictive regressions are likely to have persistent time series characteristics. The literature has sought to find a realistic approach to allow for this general phenomena. One approach that has received much attention is to develop asymptotics for inference using a local to unity autoregressive specification for the regressor x_t in (2.1) so that $\rho = 1 + \frac{c}{n}$ in (2.2) e.g., Campbell and Yogo (2006) and Jansson and Moreira (2006).

In this local to unity case, the well known "finite sample bias" in estimation is still present in the limit and correcting the asymptotic bias is generally not possible since the bias depends on the localizing coefficient c and this parameter is not consistently estimable (see below for further discussion).

Moreover, the limit theory in this case is nonstandard by virtue of the stronger signal/noise ratio and near unit root behavior in the regressor. In particular, standard methods and notation (Phillips, 1987) lead to the following limit theory

$$n\left(\hat{\beta}-\beta\right) = \frac{\frac{1}{n}\sum_{t=1}^{n} x_{t-1}u_{0t}}{\frac{1}{n^2}\sum_{t=1}^{n}(x_{t-1})^2} \Longrightarrow \frac{\int J_x^c(r)dB_0(r)}{\int J_x^c(r)^2 dr} \tag{2.6}$$

where B_0 is Brownian motion and $J_x^c(r)$ is a linear diffusion. For details on notation, conditions, and limit theory see the Appendix A.1. The limit distribution (2.6) is not mixed normal and is not pivotal. The main source of nonnormality comes from the dependence between the martingale components that are involved in the numerator and denominator together with the nonstationary regressor. More specifically the martingale $\sum_{s=1}^{t}\xi_{ns}$ where $\xi_{nt} := \left(\frac{1}{n}x_{t-1}u_{0t}, \frac{1}{\sqrt{n}}u_{xt}\right)'$ has conditional variance

$$\sum_{t=1}^{n} E_{\mathcal{F}_{nt-1}}\xi_{nt}\xi_{nt}' = \begin{bmatrix} \left(\frac{1}{n^2}\sum x_{t-1}^2\right)\Sigma_{00} & \left(\frac{1}{n\sqrt{n}}\sum x_{t-1}\right)\Sigma_{0x} \\ \left(\frac{1}{n\sqrt{n}}\sum x_{t-1}\right)\Sigma_{x0} & \Sigma_{xx}\end{bmatrix},$$

which is not diagonal unless $\Sigma_{0x} \neq 0$. The earlier decomposition (2.3) leads to

$$n\left(\hat{\beta}-\beta\right) = \frac{\frac{1}{n}\sum_{t=1}^{n} x_{t-1}u_{0.xt}}{\frac{1}{n^2}\sum_{t=1}^{n}(x_{t-1})^2} + \left(\frac{\Sigma_{x0}}{\Sigma_{xx}}\right)n\left(\hat{\rho}-\rho\right)$$

$$= \frac{\frac{1}{n}\sum_{t=1}^{n} x_{t-1}u_{0.xt}}{\frac{1}{n^2}\sum_{t=1}^{n}(x_{t-1})^2} + \left(\frac{\Sigma_{x0}}{\Sigma_{xx}}\right)\frac{\frac{1}{n}\sum_{t=1}^{n} x_{t-1}u_{xt}}{\frac{1}{n^2}\sum_{t=1}^{n}(x_{t-1})^2}. \tag{2.7}$$

Note that, with $\Sigma_{00.x} = \Sigma_{00} - \Sigma_{0x}\Sigma_{xx}^{-1}\Sigma_{x0} = \Sigma_{00} - \frac{\Sigma_{x0}^2}{\Sigma_{xx}}, \frac{\Sigma_{00.x}}{\Sigma_{00}} = 1 - \frac{\Sigma_{x0}^2}{\Sigma_{00}\Sigma_{xx}}$, and $\xi_{nt} := \left(\frac{1}{n}x_{t-1}u_{0.xt}, \frac{1}{\sqrt{n}}u_{xt}\right)'$ we have the diagonal conditional variance

$$\sum_{t=1}^{n} E_{\mathcal{F}_{nt-1}}\xi_{nt}\xi_{nt}' = \begin{bmatrix}\left(\frac{1}{n^2}\sum x_{t-1}^2\right)\Sigma_{00.x} & 0 \\ 0 & \Sigma_{xx}\end{bmatrix}$$

and the first term of (2.7) converges to the mixed normal (MN) limit,

$$\frac{\frac{1}{n}\sum_{t=1}^{n} x_{t-1}u_{0.xt}}{\frac{1}{n^2}\sum_{t=1}^{n}(x_{t-1})^2} \Longrightarrow MN\left(0, \Sigma_{00.x}\left(\int J_x^c(r)^2 dr\right)^{-1}\right),$$

while the second term is the standard unit root statistic

$$\frac{\frac{1}{n}\sum_{t=1}^{n} x_{t-1}u_{xt}}{\frac{1}{n^2}\sum_{t=1}^{n}(x_{t-1})^2} \Longrightarrow \frac{\int J_x^c(r)dB_x(r)}{\int J_x^c(r)^2 dr}.$$

Hence

$$n\left(\hat{\beta}-\beta\right) \Longrightarrow MN\left(0, \Sigma_{00.x}\left(\int J_c^x(r)^2 dr\right)^{-1}\right) + \left(\frac{\Sigma_{x0}}{\Sigma_{xx}}\right)\frac{\int J_x^c(r)dB_x(r)}{\int J_x^c(r)^2 dr}.$$

Therefore the source of the nonstandard limit distribution comes from the regression endogeneity and the persistent regressor. In fact, we can show that after using a standard fully modified endogeneity bias correction method, such as Phillips and Hansen (1990), which is designed for the unit root case ($c = 0$), we have

$$n\left(\hat{\beta}_{FM}-\beta\right) \Longrightarrow MN\left(c\left(\frac{\Sigma_{x0}}{\Sigma_{xx}}\right), \Sigma_{00.x}\left(\int J_x^c(r)^2 dr\right)^{-1}\right).$$

Thus, without precise information on c, the bias is not correctable.

Since

$$n\widehat{\sigma}_{\beta} = \left(\hat{\Sigma}_{00}\right)^{1/2}\left(\frac{1}{n^2}\sum_{t=1}^{n}(x_{t-1})^2\right)^{-1/2} \Longrightarrow \left(\Sigma_{00}\right)^{1/2}\left(\int J_x^c(r)^2 dr\right)^{-1/2},$$

$$\frac{\hat{\beta}-\beta}{\widehat{\sigma}_{\beta}} \Longrightarrow \frac{MN\left(0, \Sigma_{00.x}\left(\int J_x^c(r)^2 dr\right)^{-1}\right) + \left(\frac{\Sigma_{x0}}{\Sigma_{xx}}\right)\left(\frac{\int J_x^c(r)dB_x(r)}{\int J_x^c(r)^2 dr}\right)}{\left(\Sigma_{00}\right)^{1/2}\left(\int J_x^c(r)^2 dr\right)^{-1/2}}$$

$$= \left(\frac{\Sigma_{00.x}}{\Sigma_{00}}\right)^{1/2} Z + \left(\frac{\Sigma_{x0}}{\left(\Sigma_{xx}\Sigma_{00}\right)^{1/2}}\right)\left(\frac{\int J_x^c(r)dB_x(r)}{\left(\Sigma_{xx}\int J_x^c(r)^2 dr\right)^{1/2}}\right)$$

$$=: \phi\widehat{\tau}_{DF} + \left(1-\phi^2\right)^{1/2} Z,$$

where

$\phi = \frac{\Sigma_{x0}}{\left(\Sigma_{xx}\Sigma_{00}\right)^{1/2}}$ and $Z \sim N(0,1)$ independent of $\widehat{\tau}_{DF}$.

This result is well known (e.g. Elliott and Stock, 1994) and is used frequently in the literature (e.g. Cavanagh et al., 1995; Campbell and Yogo, 2006). Therefore, unless $\Sigma_{x0} \neq 0$, standard t-ratio testing or chi-square inference is unavailable. The source of the uncorrectable bias difficulty lies clearly in the nuisance parameter in the limit distribution and the lack of mixed normality.

2.2. Suggested solutions

A primary desirable characteristic in a solution that maintains a local to unity $\left(\rho = 1 + \frac{c}{n}\right)$ condition of persistence in x_t is a pivotal limit distribution or distribution with readily correctable bias. There are many existing and ongoing studies that seek to resolve these problems and the following is a brief summary of the main approaches.

2.2.1. The Bonferroni method

The limit distribution of the t statistic

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta}{\widehat{\sigma}_\beta} \Longrightarrow \phi \widehat{\tau}_{DF}(c) + \left(1 - \phi^2\right)^{1/2} Z$$

depends on c, which is not consistently estimable. So Cavanagh et al. (1995) introduced a pretest for identifying conditions under which the conventional t-test is approximately valid. If the conditions do not hold, they suggested a Bonferroni approach which searches possible values for c (and hence ρ) and uses the most conservative ones in constructing a confidence interval or test. To construct a Bonferroni confidence interval (CI), the investigator first constructs a $100(1-\alpha_1)$ % CI for c, denoted as $CI_c(\alpha_1)$. Then, for each value of c in this confidence interval, construct a $100(1-\alpha_2)$ % CI for β given that value of c, denoted as $CI_{\beta|c}(\alpha_2)$. A CI that does not depend on c is then obtained as

$$CI_\beta(\alpha) = \bigcup_{c \in CI_c(\alpha_1)} CI_{\beta|c}(\alpha_2).$$

By Bonferroni's inequality, the CI has coverage probability of at least $100(1-\alpha_1-\alpha_2)$ %.

More specifically, based on the estimator $\hat{\rho}$ and using the unit root t-statistic, the proposal is to find $CI_c(\alpha_1) = [c_l(\alpha_1), c_u(\alpha_1)]$ as in Stock (1991). Using the critical value $d_{t_{\hat{\beta}},c}$ of the limit variate

$$\phi \widehat{\tau}_{DF}(c) + \left(1 - \phi^2\right)^{1/2} Z,$$

the approach calculates

$$\begin{aligned} CI_\beta(\alpha_1, \alpha_2) &= \left[d_l^\beta(\alpha_1, \alpha_2), d_u^\beta(\alpha_1, \alpha_2) \right] \\ &= \left[\min_{c_l \le c \le c_u} d_{t_{\hat{\beta}},c,\frac{1}{2}\alpha_2}, \max_{c_l \le c \le c_u} d_{t_{\hat{\beta}},c,1-\frac{1}{2}\alpha_2} \right]. \end{aligned} \tag{2.8}$$

Finally, a CI for β is proposed

$$\left[\hat{\beta} - \widehat{\sigma}_\beta d_u^\beta(\alpha_1, \alpha_2), \hat{\beta} - \widehat{\sigma}_\beta d_l^\beta(\alpha_1, \alpha_2) \right],$$

for which the limit theory is

$$\begin{aligned} &\Pr\left(t_{\hat{\beta}} \notin \left[d_l^\beta(\alpha_1, \alpha_2), d_u^\beta(\alpha_1, \alpha_2) \right] \right) \\ &\to \Pr\left(\phi \widehat{\tau}_{DF}(c) + \left(1-\phi^2\right)^{1/2} Z \notin \left[d_l^\beta(\alpha_1, \alpha_2), d_u^\beta(\alpha_1, \alpha_2) \right] \right) \\ &\le \alpha_1 + \alpha_2. \end{aligned}$$

Campbell and Yogo (2006) utilized this idea by employing an augmented regression equation as in Phillips and Hansen (1990). They used the Bonferroni method above in conjunction with a DF-GLS unit root test statistic (Elliott et al., 1996) to remove the dependence on c in the confidence interval.

With a regressor whose autoregressive root is very close to unity, this approach shows successful size control while maintaining local power. The method has been frequently employed in the applied literature, but has some undesirable properties that should be noted. First, empirical size may be substantially lower than nominal size resulting in a conservative test whose power is often negligible in near local alternatives to the null of non predictability. Another critical limitation is the difficulty of extending this approach to multivariate regressions involving several predictors which induce many unknown c coefficients, substantially complicating constructions of confidence intervals of the type (2.8). Finally, Stock's confidence intervals for ρ are now known to be invalid and seriously biased asymptotically when $c \to -\infty$ (Phillips, 2012b).[2] This failure in the approach leads to poor performance in predictive regression tests based on Bonferroni methods such as those in Campbell and Yogo (2006) and Cavanagh et al. (1995) when the regressor is stationary or mildly integrated (as in (I1) below).

[2] Asymptotic analysis of the invalidity in this construction and confirmatory finite sample simulations are given in Phillips (2012b). Lee (2012, Section 4) also reports simulation results confirming the problem in predictive regression settings.

2.2.2. A conditional likelihood approach with sufficient statistics

Jansson and Moreira (2006) suggested a conditional likelihood method that uses sufficient statistics. The central idea is to find the sufficient statistics $\left(R_\beta, R_\rho, R_{\beta\beta}, R_{\rho\rho}\right)$ for (β, ρ) in (2.1) and (2.2). A test for β is then constructed from the conditional likelihood of $\left(R_\beta, R_\rho\right)$ given $\left(R_{\beta\beta}, R_{\rho\rho}\right)$, whose distribution does not depend on β. Final critical value functions are obtained from the conditional likelihood of R_β given $\left(R_\rho, R_{\beta\beta}, R_{\rho\rho}\right)$. The test based on this approach attains conditional optimality within a certain class, and has therefore also received attention. Like the Bonferroni method, the approach also has some undesirable aspects. In particular, it is difficult to extend beyond a single regressor model — for the same reason as before — and the algorithm for implementation involves some highly complicated numerical quadrature that is known to present numerical difficulties in implementation (Kasparis et al., 2012). Both properties reduce the appeal of this method for applied research.

2.2.3. A control function approach

Recent work by Elliott (2011) proposed adding a stationary variable to the predictive regression to help stabilize the limit theory. The idea stems from the following augmented system assuming a known ρ

$$\begin{aligned} y_t &= \beta x_{t-1} + u_{0t} = \beta x_{t-1} + \frac{\Sigma_{0x}}{\Sigma_{xx}} u_{xt} + u_{0.xt} \\ &= \beta x_{t-1} + \frac{\Sigma_{0x}}{\Sigma_{xx}} (1 - \rho L) x_t + u_{0.xt}. \end{aligned}$$

By the same logic as before we have a mixed normal limit theory for $\hat{\beta}$ regardless of ρ. But since this procedure is not feasible Elliott (2011) suggested finding a proxy orthogonalizing variable z_t leading to

$$y_t = \beta x_{t-1} + \alpha z_t + \tilde{u}_{0t}$$

so that the correlation between $\tilde{u}_{0t}$ and u_{xt} is less than that of u_{0t} and u_{xt}, thereby diminishing the effect of endogeneity in the regression. Then the following limit theory applies

$$\frac{\hat{\beta} - \beta}{\widehat{\sigma}_\beta} \Longrightarrow \tilde{\phi} \widehat{\tau}_{DF} + \left(1 - \tilde{\phi}^2\right)^{1/2} Z,$$

with $|\tilde{\phi}| < |\phi|$. In simulations this approach was shown to have better size control with higher local power than the infeasible Campbell and Yogo method (based on a known value of ρ) in the presence of perfect orthogonalizing regressors that might be suggested by economic theory. However, in the absence of a perfect orthogonalizing variable, the approach cannot completely remove the nonstandard and non pivotal features of the limit distribution.

2.2.4. The IVX approach

Another recent approach to predictive regression relies on the IVX method of MP. The idea of IVX is to generate a less persistent instrument for the regressor than the regressor itself which uses no extraneous information so that the instrument relies only on the regressor. The instrument construction takes the form

$$\tilde{z}_t = \sum_{j=1}^{t} \rho_{nz}^{t-j} \triangle x_j$$

$$\rho_{nz} = 1 + \frac{c_z}{n^\delta}, \quad \delta \in (0, 1), c_z < 0,$$

which is clearly dependent only on $\{x_t\}$ - hence the terminology IVX – and parameters $\delta \in (0, 1)$ and $c_z < 0$ which are specified by the investigator to ensure that $\tilde{z}_t$ is less persistent than the regressor x_t. Since $\triangle x_j = \frac{c}{n} x_{j-1} + u_{xj}$,

$$\tilde{z}_t = \sum_{j=1}^{t} \rho_{nz}^{t-j} \left(\frac{c}{n} x_{j-1} + u_{xj} \right) = \sum_{j=1}^{t} \rho_{nz}^{t-j} u_{xj} + \frac{c}{n} \sum_{j=1}^{t} \rho_{nz}^{t-j} x_{j-1}$$
$$= z_t + \frac{c}{n} \psi_{nt},$$

and $z_t = \rho_{nz} z_{t-1} + u_{xt}$ plays the role of mildly integrated instrument. The remainder $\frac{c}{n} \psi_{nt}$ turns out to be asymptotically negligible, which enables nuisance parameter free inference. Then, it can be shown that the IVX estimator

$$\hat{\beta}_{IVX} = \frac{\sum_{t=1}^{n} \tilde{z}_{t-1} y_t}{\sum_{t=1}^{n} \tilde{z}_{t-1} x_{t-1}} = \beta + \frac{\sum_{t=1}^{n} \tilde{z}_{t-1} u_{0t}}{\sum_{t=1}^{n} \tilde{z}_{t-1} x_{t-1}}$$

has a following limit theory,

$$n^{\frac{1+\delta}{2}} \left(\hat{\beta}_{IVX} - \beta \right) \Longrightarrow \psi',$$

where ψ' is a correctly centered mixed normal random variable, and

$$\frac{\hat{\beta}_{IVX} - \beta}{\widehat{\sigma}_{IVX}} \Longrightarrow Z,$$

which is standard normal. The key element in this limit theory is the asymptotic independence between the martingale part of $\sum_{t=1}^{n} \tilde{z}_{t-1} u_{0t}$ and $\sum_{t=1}^{n} \tilde{z}_{t-1} x_{t-1}$, which obtains by virtue of the reduced order of magnitude of z_t (hence $\tilde{z}_t$). Compared to the earlier discussion, we now have the mds $\xi_{nt}' := (\frac{1}{n^{\frac{1+\delta}{2}}} z_{t-1} u_{0t}, \frac{1}{\sqrt{n}} u_{xt})'$ and martingale conditional variance

$$\sum_{t=1}^{n} \mathbb{E}_{\mathcal{F}_{nt-1}} \xi_{nt} \xi_{nt}'$$
$$= \begin{bmatrix} \left(\frac{1}{n^{1+\delta}} \sum z_{t-1}^2 \right) \Sigma_{00} & \left(\frac{1}{n^{1+\frac{\delta}{2}}} \sum z_{t-1} \right) \Sigma_{0x} \\ \left(\frac{1}{n^{1+\frac{\delta}{2}}} \sum z_{t-1} \right) \Sigma_{0x} & \Sigma_{xx} \end{bmatrix},$$

which has a diagonal limit since $\left(\frac{1}{n^{1+\delta}} \sum_{t=1}^{n} z_{t-1}^2 \right) = O_p(1)$ but $\sum_{t=1}^{n} z_{t-1} = O_p(n^{\frac{1}{2}+\delta})$ for mildly integrated z_t as shown in Phillips and Magdalinos (2007a,b), so $n^{-(1+\frac{\delta}{2})} \sum_{t=1}^{n} z_{t-1} = n^{-(\frac{1-\delta}{2})} \times O_p(1) = o_p(1)$.

This approach to predictive regression has many desirable properties. It leads to a pivotal mixed normal limit distribution under persistent regressors, and the degree of allowable persistence in the regressor is quite general, including mildly integrated, local to unity and unit root regressors. In another paper by the same authors (Phillips and Lee, 2012b), we show that even under the locally and mildly explosive regressors, the chi-square limit theory remains robust. Another attractive feature of the framework is that it is very straightforward to extend to multivariate systems. Thus, most of the existing difficulties of inference in predictive regressions are nicely resolved by this approach. Kostakis et al. (2010) and Gonzalo and Pitarakis (2009) have applied an IVX approach to the predictive regression setting.

3. Long-horizon IVX

Studies on short horizon predictive regression frequently found no significant predictive power or at most a marginal degree of explanatory power. Long-horizon regression was proposed as an alternate model and has been extensively used to argue the predictability of stock returns over a reasonable time horizon. The process of aggregation inevitably lends some persistence to a noisy dependent variable since the time sum of any $I(0)$ variable evidently becomes more persistent as the horizon (or period of summation) increases. Currently, the most popular econometric procedure is Valkanov (2003). Under the assumption of an increasing horizon $k = O(n)$, the author suggested an asymptotically valid procedure based on a well defined limit distribution of the test. Nuisance parameters were handled as in Cavanagh et al. (1995) and Campbell and Yogo (2006) in conjunction with another procedure. Accordingly, the distributions of the test statistics are again nonstandard and simulations are needed to compute critical values. The procedure is also restricted to the scalar regressor case. So the approach has similar limitations to many of those described earlier.

As noted above, the IVX approach addresses these limitations. In view of the promise in this approach, the present Section develops a long-horizon version of IVX and shows that robust inference with standard asymptotics is available using this method. In the analytic framework of long-horizon regression, specifications for both a fixed and increasing horizon k have been used where in the latter case k is proportional to the sample size n. We focus here on a bridging intermediate case where k may increase according to the condition $\frac{k}{n} + \frac{1}{k} \to 0$. This framework is general enough to cover most cases of practical interest involving long horizon forecasting.

3.1. Model framework

We consider the multivariate predictive regression system

$$y_{t+1} = A x_t + u_{0t+1}, \tag{3.1}$$
$$x_{t+1} = R_n x_t + u_{xt+1},$$
$$R_n = I_p + \frac{C}{n^{\alpha}}, \quad \text{for some } \alpha > 0,$$

where A is an $m \times p$ coefficient matrix and $C = diag\left(c_1, c_2, \ldots, c_p\right)$ is a diagonal matrix of localizing coefficients that are unknown and provide some flexibility in the properties of the multivariate regressors. We allow for more general degrees of persistence in the regressors than the existing literature since x_t may belong to any of the following categories:

(I1) mildly integrated ($C < 0, \alpha \in (0, 1)$),
(I2) near integrated ($C < 0, \alpha = 1$),
(I3) integrated ($C = 0$),
(I4) locally explosive ($C > 0, \alpha = 1$),
(I5) mildly explosive ($C > 0, \alpha \in (0, 1)$).

This framework adds considerable generality that is useful in both empirical work and theory.[3] Existing studies of predictive regression typically considered only the near integrated scalar regressor case and relied on asymptotic results for such processes (such as Phillips, 1987; Chan and Wei, 1987) combined with Bonferroni methods to control test size. One reason for the restriction to near integrated processes as regressors is that the asymptotic theory for time series with a local to unity parameter has long been well known among empirical researchers and is better understood than the limit theory in wider vicinities of unity. The recent work on limit theory encompassing a wider autoregressive parameter

[3] Further flexibility may be introduced by allowing variation in the power exponent rate parameter α across regressors, thereby directly influencing the degree of persistence in individual regressors. This additional level of generality is not considered in the present work but may be handled by the methods given here and in MP.

space (Phillips and Magdalinos, 2007a,b; Magdalinos and Phillips, 2009a) enables a more generally applicable approach to predictive regression. That approach is pursued here.

As noted, a further advantage of the approach we consider is that all cases (I1)–(I5) above are encompassed in the theory in the context of multivariate regressors whose autoregressive roots can lie in a wide vicinity of unity. The approach is therefore compatible with models where there are a variety of economic fundamentals or factors that serve as common elements in determining variables such as stock returns or the forward premium, as is common in empirical finance models (Fama and French, 1993; Kothari and Shanken, 1997; Lewellen, 2004, and many others).

To proceed, we formulate a general weakly dependent innovation structure of the linear process form

$$u_t := \begin{bmatrix} u_{0t} \\ u_{xt} \end{bmatrix} = \sum_{j=0}^{\infty} F_j \varepsilon_{t-j}, \qquad \varepsilon_t = \begin{bmatrix} \varepsilon_{0t} \\ \varepsilon_{xt} \end{bmatrix} \sim mds\ (0, \Sigma),\ \Sigma > 0, \qquad \mathbb{E}\|\varepsilon_1\|^4 < \infty, \tag{3.2}$$

$$F_0 = I_{m+p}, \qquad \sum_{j=0}^{\infty} j\|F_j\| < \infty, \qquad F(z) = \sum_{j=0}^{\infty} F_j z^j \quad \text{and} \quad F(1) = \sum_{j=0}^{\infty} F_j > 0. \tag{3.3}$$

Following convention in the predictive regression literature, we impose an mds structure on the regression error u_{0t} using the following restricted F_j matrices

$$F_j = \begin{bmatrix} F_{0j} \\ F_{xj} \end{bmatrix}, \qquad F_{0j} = \begin{cases} [I_m : 0_{m\times p}] & \text{for } j = 0 \\ 0_{p\times(m+p)} & \text{for } j \geq 1 \end{cases}, \tag{3.4}$$

giving a special case of (3.2) and (3.3). This formulation allows general linear dependence in u_{xt} but imposes an mds structure on u_{0t} to ensure that there is non-predictability of y_t under the null hypothesis, as is explained below in Section 3.2. While this case is of great practical relevance in some financial applications, dependent u_{0t} specifications are also meaningful, sometimes in the same context. For example, a bivariate regression specification with mds innovations as in (2.1) is unconvincing when there are two or more significant predictors. To see this, let x_{1t} and x_{2t} be the dividend–price and earnings price ratios, two explanatory variables for stock returns for which many authors have found significant predictive power (e.g. Campbell and Yogo, 2006). Then, taking a single regressor specification (as is commonly done in empirical research such as Campbell and Yogo) we have

$$y_{t+1} = \beta x_{1t} + u_{0t+1},$$
$$\mathbb{E}(u_{0t+1}|\mathcal{F}_t) \neq 0 \quad \text{since } x_{2t} \in \mathcal{F}_t,$$

thereby contradicting the mds condition.

On the other hand, even though we impose (3.4), the multiple predictor model is less subject to omitted variable misspecification of this type. Of course, the mds assumption also conforms with non-predictability under the null in Section 3.2. In what follows, the notation we use can be interpreted as belonging to the special mds case. For example, the matrices Ω_{0x} and Ω_{00} are simply special cases of general long-run covariances corresponding to contemporaneous covariances.

We denote the long run covariance matrices associated with u_t as

$$\Omega = \sum_{h=-\infty}^{\infty} \mathbb{E}\left(u_t u_{t-h}'\right) = F(1)\Sigma F(1)',$$
$$F(1) = \begin{bmatrix} F_0(1) \\ F_x(1) \end{bmatrix} = \begin{bmatrix} [I_m : 0_{m\times p}] \\ F_x(1) \end{bmatrix},$$

$$\Lambda = \sum_{h=1}^{\infty} \mathbb{E}\left(u_t u_{t-h}'\right), \qquad \Delta = \Lambda + \mathbb{E}\left(u_1 u_1'\right),$$
$$\Omega = \begin{bmatrix} \Omega_{00} & \Omega_{0x} \\ \Omega_{x0} & \Omega_{xx} \end{bmatrix}, \qquad \Lambda = \begin{bmatrix} \Lambda_{00} & \Lambda_{0x} \\ \Lambda_{x0} & \Lambda_{xx} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_{xx} \end{bmatrix},$$

and use the functional law (Phillips and Solo, 1992)

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor ns \rfloor} u_j = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor ns \rfloor} \begin{bmatrix} u_{0t} \\ u_{xt} \end{bmatrix} = \begin{bmatrix} B_{0n}(s) \\ B_{xn}(s) \end{bmatrix} \Longrightarrow \begin{bmatrix} B_0(s) \\ B_x(s) \end{bmatrix} = BM \begin{bmatrix} \Omega_{00} & \Omega_{0x} \\ \Omega_{x0} & \Omega_{xx} \end{bmatrix},$$

and local to unity limit law for cases (I2)–(I4) (Phillips, 1987):

$$\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} \Longrightarrow J_x^c(r), \tag{3.5}$$

where $J_x^c(r) = \int_0^r e^{(r-s)C} dB_x(s)$, which encompasses the unit root case where $J_x^c(r) = B_x(r)$ when $C = 0$. Under (3.2)–(3.3) we have the Beveridge–Nelson (BN) decomposition (Phillips and Solo, 1992)

$$u_t = F(1)\varepsilon_t - \Delta\tilde{\varepsilon}_t, \quad \tilde{\varepsilon}_t = \sum_{j=0}^{\infty} \tilde{F}_j \varepsilon_{t-j}, \ \tilde{F}_j = \sum_{s=j+1}^{\infty} F_s.$$

Note that $\tilde{F}_{0j} = 0$ and hence $\tilde{\varepsilon}_{0t} = 0$ for all j and t, from (3.4). The component decompositions are then

$$u_{0t} = F_0(1)\varepsilon_t - \Delta\tilde{\varepsilon}_{0t} = \varepsilon_{0t},$$
$$u_{xt} = F_x(1)\varepsilon_t - \Delta\tilde{\varepsilon}_{xt}.$$

3.2. Rearranging the regression and long-horizon IVX

Because of its flexibility and analytic convenience we consider a moderately increasing horizon such as $k = n^\nu$, with $\nu \in (0, 1)$, and accordingly impose the horizon rate condition

$$\frac{1}{k} + \frac{k}{n} \to 0,$$

which has substantial generality. Application of the IVX procedure in its original form leads to inconsistent estimation, just as in Valkanov (2003). However, a simple rearrangement of the regression about the null hypothesis is useful in producing a stronger regressor signal and consistent estimation. Such a rearrangement was originally suggested in Jegadeesh (1991) and Cochrane (1991) for stationary regressor cases in order to vitiate the effects of serial correlation of the residuals in long-horizon regression. Liu and Maynard (2007) recently used this same rearrangement to make the dependent variable white noise under the null and employed sign and signed rank regression methods. The same idea is applied here to empirical long-horizon regressions.

The starting point is to consider testing predictability in a long horizon regression based on the following empirical regression and null hypothesis:

$$y_{t+1}^k = Ax_t + u_{0t+1}^k \tag{3.6}$$

$$\mathcal{H}_0(k) : \mathbb{E}_t\left[y_{t+1}^k\right] = 0 \quad \text{or} \quad A = 0, \tag{3.7}$$

$$y_{t+1}^k = \sum_{j=1}^{k} y_{t+j}, \qquad u_{0t+1}^k = \sum_{j=1}^{k} u_{0t+j},$$

where we use the natural filtration $\mathcal{F}_t = \sigma(x_0, y_0, u_t, u_{t-1}, \ldots)$ with $u_t = (u_{0t}', u_{xt}')'$ and set $\mathbb{E}_{\mathcal{F}_t}(\cdot) = \mathbb{E}_t(\cdot)$. The alternative hypothesis is $\mathcal{H}_1(k) : \mathbb{E}_t\left[y_{t+1}^k\right] \neq 0$ or $A \neq 0$. This type of empirical regression has been frequently used in the literature. As discussed in Hjalmarsson (2011), Eq. (3.6) is formulated not as the data-generating process (DGP) but for a fitted regression. The conventional DGP employed in the literature describes the mechanism $y_{t+1} = Ax_t + u_{0t+1}$ and $x_{t+1} = R_n x_t + u_{xt+1}$ as in (3.1). So y_t is an mds with the natural filtration $\mathcal{F}_t$ under the null of non-predictability in this DGP. One statistical obstacle in the model (3.6) is that the innovation u^k_{0t+1} is a partial sum of $I(0)$ variables and therefore has some persistence characteristics which affect the estimation limit theory, including that of IVX.

We therefore seek an alternative empirical regression for testing predictability or absence of predictability in long horizon regression. To motivate the construction we note that if

$$\mathbb{E}_t[y_{t+k}] = \mathbb{E}_{t+1}[y_{t+k}] = \cdots = \mathbb{E}_{t+k-1}[y_{t+k}] = 0, \tag{3.8}$$

then in view of stationarity of y_{t+k} and under the null of unpredictability

$$\mathbb{E}_{t+1}[y_{t+k}] = 0 \Rightarrow^{(\text{1-period backshifting})} \mathbb{E}_t[y_{t+k-1}] = 0,$$
$$\mathbb{E}_{t+2}[y_{t+k}] = 0 \Rightarrow^{(\text{2-period backshifting})} \mathbb{E}_t[y_{t+k-2}] = 0,$$
$$\vdots$$
$$\mathbb{E}_{t+k-1}[y_{t+k}] = 0 \Rightarrow^{((k-1)\text{-period backshifting})} \mathbb{E}_t[y_{t+1}] = 0.$$

It follows that:

$$\mathbb{E}_t[y_{t+k}] = \mathbb{E}_t[y_{t+k-1}] = \cdots = \mathbb{E}_t[y_{t+1}] = 0. \tag{3.9}$$

This formulation motivates regressing one period returns on a long-horizon version of the regressor, leading to the following empirical regression:

$$y_{t+k} = Bx^k_t + u_{0t+k}, \qquad x^k_t = \sum_{j=1}^{k} x_{t+j-1} \tag{3.10}$$

$$\mathcal{H}'_0(k) : B = 0. \tag{3.11}$$

for which the alternative hypothesis is $\mathcal{H}'_1(k) : B \neq 0$. Later in the paper we will consider model validity under a class of local alternatives within $\mathcal{H}'_1(k)$. Note that (3.8), which is equivalent to $\mathcal{H}'_0(k)$, implies (3.9) and hence eventually guarantees $\mathcal{H}_0(k)$, i.e.,

$$\mathcal{H}'_0(k) : B = 0 \Leftrightarrow \left\{ \begin{array}{c} \mathbb{E}_t[y_{t+k}] = 0 \\ \mathbb{E}_{t+1}[y_{t+k}] = 0 \\ \vdots \\ \mathbb{E}_{t+k-2}[y_{t+k}] = 0 \\ \mathbb{E}_{t+k-1}[y_{t+k}] = 0 \end{array} \right\}$$

$$\Rightarrow \left\{ \begin{array}{c} \mathbb{E}_t[y_{t+1}] = 0 \\ \mathbb{E}_t[y_{t+2}] = 0 \\ \vdots \\ \mathbb{E}_t[y_{t+k-1}] = 0 \\ \mathbb{E}_t[y_{t+k}] = 0 \end{array} \right\} \Rightarrow \mathcal{H}_0(k) : A = 0.$$

The first equivalence ($\Leftrightarrow$) above holds since the empirical regression coefficient B in (3.10) satisfies the following relation for $j = 1, \ldots, k$,

$$\mathbb{E}_{t+j-1}[y_{t+k}] = Bx^j_t + B\mathbb{E}_{t+j-1}\left[x^{k-j}_{t+j}\right],$$

so that $B = 0$ implies $\mathbb{E}_{t+j-1}[y_{t+k}] = 0$ for all $j = 1, \ldots, k$. The opposite direction is easy to show by contraposition but for our purposes here we only need the primary direction. In particular, the alternative empirical regression (3.10) produces a test that provides a sufficient condition for the test in the original regression (3.6) in the sense that $\mathcal{H}'_0(k)$ implies $\mathcal{H}_0(k)$.

Under the alternative, the two models (3.6) and (3.10) have different specifications. In Appendix A.2, we relate local alternatives to $\mathcal{H}_0(k)$ based on (3.1) and those of $\mathcal{H}'_0(k)$ from (3.10). We then analyze the local asymptotic power properties of the long horizon IVX procedure suggested below. Using local alternatives, Appendix A.2 also provides a mechanism for rectifying the apparently unbalanced nature of the regression, which has been a universal problem in the predictive regression literature, e.g., see the discussions in Gospodinov (2009) or Torous and Valkanov (2000). The remainder of the main text develops inference procedures and asymptotics under the maintained hypothesis of non-predictability since the focus in predictive regression testing is still primarily on the behavior of the tests under the null hypothesis.

In the model (3.10) the regressors are persistent and the innovations are $I(0)$, so the equation's stronger signal ensures regression consistency. Importantly with this formulation, a modified IVX procedure is applicable that removes dependence of the test statistics on the localizing coefficient nuisance parameter C in x_t and enables us to enjoy the benefits of IVX regression.

In short-horizon predictive regression, the IVX instrument is constructed as

$$\tilde{z}_t = \sum_{j=1}^{t} R^{t-j}_{nz} \triangle x_j,$$

$$\text{with } R_{nz} = I_p - \frac{C_z}{n^\delta}, \quad \delta \in (0,1), C_z = c_z I_p \text{ and } c_z > 0.$$

Since $\triangle x_j = \frac{C}{n^\alpha} x_{j-1} + u_{xj}$,

$$\tilde{z}_t = \sum_{j=1}^{t} R^{t-j}_{nz} \left(\frac{C}{n^\alpha} x_{j-1} + u_{xj} \right)$$

$$= \sum_{j=1}^{t} R^{t-j}_{nz} u_{xj} + \frac{C}{n^\alpha} \sum_{j=1}^{t} R^{t-j}_{nz} x_{j-1} \tag{3.12}$$

$$= z_t + \frac{C}{n^\alpha} \psi_{nt}. \tag{3.13}$$

After decomposition, $z_t = R_{nz} z_{t-1} + u_{xt}$ plays the role of a mildly integrated instrument and the remainder ψ_{nt} is controllable due to its coefficient $\frac{C}{n^\alpha}$ (except for the mildly explosive (I5) case — see (Phillips and Lee, 2012b)), which leads to robust, nuisance parameter free inference.

Since the long-horizon regressor in (3.10) is $x^k_t = \sum_{j=1}^{k} x_{t+j-1}$, we design a long-horizon IVX (LHIVX) approach using the following IVX instruments

$$\tilde{z}^k_t = \sum_{j=1}^{k} \tilde{z}_{t+j-1}, \quad \tilde{z}_t = \sum_{j=1}^{t} R^{t-j}_{nz} \triangle x_j$$

$$R_{nz} = I_p - \frac{C_z}{n^\delta}, \quad \delta \in (0,1), C_z > 0.$$

Then

$$\tilde{z}^k_t = z^k_t + \frac{C}{n^\alpha} \psi^k_{nt}, \tag{3.14}$$

where

$$z^k_t = \sum_{j=1}^{k} z_{t+j-1}, \quad z_t = \sum_{j=1}^{t} R^{t-j}_{nz} \left(\triangle x_j - \frac{C}{n^\alpha} x_{j-1} \right)$$

$$= \sum_{j=1}^{t} R^{t-j}_{nz} \triangle_C x_j = \sum_{j=1}^{t} R^{t-j}_{nz} u_{xj},$$

$$\psi^k_{nt} = \sum_{j=1}^{k} \psi_{nt+j-1}, \quad \psi_{nt} = \sum_{j=1}^{t} R^{t-j}_{nz} x_{j-1}.$$

We impose the rate condition

$$\frac{n^{\frac{1}{2}}}{n^{\delta}} + \frac{n^{\delta}}{k} + \frac{k}{n} \to 0,$$

so that the prediction horizon increases slower than the sample size n, but faster than the degree of mild integration in the IVX variate. The faster rate $\delta > 0.5$ indicates that the mildly integrated instrument should not be too close to stationarity, a point on which MP provide a detailed discussion.

Simulations in Phillips and Lee (2012b) show that the size performance of the IVX procedure is robust to choices of $\delta \in (0.5, 1)$ with a normalized value $C_z = -5I_p$, even for sample sizes as low as $n = 100$. Local discriminatory power also seems satisfactory with a slight tendency for power to increase with the magnitude of δ. In view of the rate condition above, we recommend for practical work that δ be chosen close to but less than the horizon rate ν in $k = n^{\nu}$. For example, if the horizon is set as $k = n^{0.75}$ then the IVX rate parameter might be chosen as $\delta = n^{0.7}$. Note that $k = n^{\nu}$ itself is chosen according to the selected length of the horizon and that larger k typically leads to higher local power (see Appendix A.2).

To test the null of unpredictability (3.11) in the long horizon regression (3.10) we propose the LHIVX estimator

$$\hat{B}^{LHIVX} = \left(\sum_{t=1}^{n-k} y_{t+k}\left(\tilde{z}_t^k\right)'\right)\left(\sum_{t=1}^{n-k} x_t^k\left(\tilde{z}_t^k\right)'\right)^{-1},$$

whose asymptotic null distribution is mixed Gaussian for a wide class of processes as will be shown in subsequent Sections. The estimator is an extended version of the IVX estimator of MP.

The following Sections develop the limit theory for testing predictive capability using LHIVX regression. We start by considering the unit root and near integrated cases and then move on to consider the mildly integrated and mildly explosive cases.

3.3. Limit theory with $\alpha = 1$

This case covers unit roots, near integration and locally explosive cases, i.e., cases (I2)–(I4). The specification is more general than the conventional local to unity specification in predictive regression which is largely preoccupied with (I2). The LHIVX estimator has estimation error under the null $B = 0$,

$$\hat{B}^{LHIVX} - B = \left(\sum_{t=1}^{n-k} u_{0t+k}\left(\tilde{z}_t^k\right)'\right)\left(\sum_{t=1}^{n-k} x_t^k\left(\tilde{z}_t^k\right)'\right)^{-1}, \tag{3.15}$$

and we provide its limit theory.

The following result gives the limit theory for $\hat{B}^{LHIVX}$ covering the cases (I2)–(I4).

Theorem 3.1. *Under the rate condition* $\frac{n^{\frac{1}{2}}}{n^{\delta}} + \frac{n^{\delta}}{k} + \frac{k}{n} \to 0$, *we have*

$$vec\left\{n^{\frac{1}{2}}k^{\frac{3}{2}}\left(\hat{B}^{LHIVX} - B\right)\right\} \Longrightarrow MN\left(0, \Sigma_B\right),$$

where

$$\Sigma_B = \left(\Psi_{cxz}^{-1}\right)' C_z^{-1}\Omega_{xx}C_z^{-1}\left(\Psi_{cxz}^{-1}\right) \otimes \Omega_{00},$$

$$\Psi_{cxz} = \frac{1}{2}\Omega_{xx}C_z^{-1} + \left(\int_0^1 J_x^c(r) J_x^c(r)' dr\right) C_z^{-1}C.$$

The limit theory is mixed normal (MN) and the convergence rate is fast. But in the unit root case I(3), as mentioned earlier, $J_x^c(r) = B_x(r)$ and $C = 0$, so the second part of Ψ_{cxz} disappears, the variance matrix Σ_B is non random, and we have a pivotal limiting normal distribution after using consistent estimators for the long-run variances.

In the general case where $C \neq 0$ the nuisance parameter dependency in the denominator can be removed simply by using a self-normalized estimator. In this way standard pivotal inference obtains. In particular, the self-normalized estimator given in the following theorem provides a convenient tool for robust inference across (I2), (I3) and (I4) cases uniformly in long-horizon prediction regression.

Theorem 3.2.

$$vec\left\{\left(\hat{B}^{LHIVX} - B\right)\right\}'\left[\left(X'P_{\tilde{Z}}X\right)^{-1} \otimes \hat{\Omega}_{00}\right]^{-1} \times vec\left\{\left(\hat{B}^{LHIVX} - B\right)\right\} \Longrightarrow \chi^2(mp),$$

where

$$\left(X'P_{\tilde{Z}}X\right)^{-1} = \left\{\left(\sum_{t=1}^{n-k} x_t^k\left(\tilde{z}_t^k\right)'\right) \times \left(\sum_{t=1}^{n-k}\left(\tilde{z}_t^k\right)\left(\tilde{z}_t^k\right)'\right)^{-1}\left(\sum_{t=1}^{n-k} x_t^k\left(\tilde{z}_t^k\right)'\right)'\right\}^{-1},$$

with any consistent estimator $\hat{\Omega}_{00}$ *for* Ω_{00}.

In view of the mds assumption (3.4) for the regression errors, a simple consistent estimator $\hat{\Omega}_{00}$ is the residual variance matrix $n^{-1}\sum_{t=1}^{n}\hat{u}_{0t}\hat{u}_{0t}'$ for any regression residuals (including OLS and IVX) for all (I1)–(I5) cases. In other applications with dependent errors u_{0t}, conventional lag kernel HAC estimators can be employed—see Kostakis et al. (2010).

Unlike earlier work on predictive regression such as Valkanov (2003), the regressors in the empirical model can be multivariate and their autoregressive roots may fall into any of the following classes: near integration (I2), unit roots (I3), or locally explosive roots (I4). Moreover, we do not need to simulate critical values since the limit theory is chi square and is free of any nuisance parameter. These advantages are not present in Bonferroni type methods even with the (I2) case alone. It turns out that these benefits to LHIVX also hold in cases of mild integration (I1) and mildly explosive roots (I5), as we now discuss.

3.4. Mildly integrated regressors

In spite of the extensive research in time series econometrics it is still challenging to discriminate clearly between stationarity and unit roots in empirical data. Beyond power and size deficiencies in conventional unit root tests, a potential explanation for some empirical results is that many economic time series have roots that are in a wide local region of unity rather than strictly at unity or in its immediate vicinity. Phillips and Magdalinos (2007a,b) and Magdalinos and Phillips (2009a) recently analyzed this type of wide vicinity of unity and called processes with this characteristic on the stationary side of unity mildly integrated processes. Many variables regarded as economic fundamentals, such as Treasury Bill rates, show varying degrees of persistence and seem to be better classified as mildly integrated processes with roots in a broad vicinity of unity rather than as strictly unit root or near integrated processes. It therefore seems worthwhile to develop asymptotics for predictive regression that allows for such behavior in the regressors. Accordingly, this Section focuses on the limit theory of the LHIVX estimator with mildly integrated regressors ($C < 0, \alpha \in (0, 1)$), i.e., case (I1) in the earlier notation.

We impose the technical restriction $\frac{k}{n^{\alpha}} \to 0$, which requires the degree of mild integration (measured by n^{α}) to exceed asymptotically the prediction horizon (k) used in the empirical regression. This condition is not too restrictive because there is flexibility in the choice of the empirical horizon k, so the findings from the predictive regression test can be regarded as robust up to this condition on the horizon. Note that although the rate parameter α is generally unknown, it is also possible to consistently estimate this parameter in practice, as shown in Phillips (2012a). Taken together with the earlier rate conditions, we therefore impose $\frac{n^{1/2}}{n^{\delta}} + \frac{n^{\delta}}{k} + \frac{k}{n^{\alpha}} + \frac{n^{\alpha}}{n} \to 0$.

We continue to work with the LHIVX estimator $\hat{B}^{LHIVX}$. The following limit theory for $\hat{B}^{LHIVX}$ applies in the mildly integrated regressor case.

Theorem 3.3.

$$vec\left\{n^{\frac{1}{2}}k^{\frac{3}{2}}\left(\hat{B}^{LHIVX} - B\right)\right\} \Longrightarrow N\left(0, 4\Omega_{xx}^{-1} \otimes \Omega_{00}\right).$$

We have the same rate of convergence as in the $\alpha = 1$ (I2)–(I4) cases, which is a similar finding to that of MP for the original IVX estimator. We also confirm that for mildly integrated regressors, the limit distribution is normal rather than mixed normal. The following theorem provides a robust chi square test that is free of nuisance parameter for cases (I1)–(I4) cases in combination with Theorem 3.2.

Theorem 3.4.

$$vec\left\{\left(\hat{B}^{LHIVX} - B\right)\right\}'\left[\left(X'P_{\tilde{Z}}X\right)^{-1} \otimes \hat{\Omega}_{00}\right]^{-1} \times Vec\left\{\left(\hat{B}^{LHIVX} - B\right)\right\} \Longrightarrow \chi^2(mp).$$

3.5. Mildly explosive regressors

This Section develops the limit theory for mildly explosive regressors $C > 0, \alpha \in (0, 1)$ covering case (I5). Booms and financial exuberance are recurring features in economic activity and these phenomena can be well characterized over subperiods by mildly explosive processes, as described in Phillips et al. (2011) and Phillips and Yu (2011). This parameter region is not covered in any of the earlier literature on predictive regression, although there are certainly subperiods of financial exuberance and rising economic fundamentals that may be modeled as temporarily explosive or very mildly so. Such periods may also have predictive content, especially in cases of contagion. Hence it is of some interest to include such cases in our analysis to achieve a comprehensive coverage in predictive regression tests.

We show below that the asymptotic behavior of the LHIVX test is robust to mildly explosive behavior in the regressors under certain conditions. In particular, we assume that $\delta < \alpha$, so that the regressor's mildly explosive behavior (measured by the parameter α) is closer to the local to unity parameter region than is the generated instrument for which $R_{nz} = I_p - \frac{C_z}{n^{\delta}}$ (see (3.16) below) and whose behavior is largely determined by δ. This condition does not seem strong as empirical evidence suggests that economic exuberance is intermittent and interspersed with more normal periods of unit root or near integrated behavior, so that explosive behavior is often only a mild departure from normality.

We define the LHIVX instruments in the same way as before, viz.,

$$\tilde{z}_t^k = \sum_{j=1}^{k} \tilde{z}_{t+j-1}, \quad \tilde{z}_t = \sum_{j=1}^{t} R_{nz}^{t-j} \triangle x_j$$

$$R_{nz} = I_p - \frac{C_z}{n^{\delta}}, \quad \delta \in (0, 1), C_z > 0. \tag{3.16}$$

Then, the decomposition

$$\tilde{z}_t^k = z_t^k + \frac{C}{n^{\alpha}} \psi_{nt}^k \tag{3.17}$$

holds as earlier with $z_t^k = \sum_{j=1}^{k} z_{t+j-1}$, $z_t = \sum_{j=1}^{t} R_{nz}^{t-j}(\triangle x_j - \frac{C}{n^{\alpha}} x_{j-1}) = \sum_{j=1}^{t} R_{nz}^{t-j} u_{xj}$, and $\psi_{nt}^k = \sum_{j=1}^{k} \psi_{nt+j-1}$, $\psi_{nt} = \sum_{j=1}^{t} R_{nz}^{t-j} x_{j-1}$. However, the remainder term ψ_{nt} plays a different role with mildly explosive regressors. The signal strength of the regressors is stronger for mildly explosive processes than for persistent regressors. So the order of magnitude of ψ_{nt} is also larger and that component of (3.17) ends up dominating the IVX asymptotics in spite of the coefficient C/n^{α} — see Phillips and Lee (2012b) for further details. Consequently, the LHIVX remainder ψ_{nt}^k also behaves differently and dominates the asymptotics just as in Phillips and Lee (2012b).

We have the following limit theory for the estimator $\hat{B}^{LHIVX}$ under mildly explosive regressors.

Theorem 3.5.

$$vec\left[kn^{\alpha}\left(\hat{B}^{LHIVX} - B\right)R_n^n\right] \Longrightarrow MN\left(0, \left(\int_0^{\infty} e^{-pC} Y_C Y_C' e^{-pC} dp\right)^{-1} \otimes \Omega_{00}\right), \tag{3.18}$$

where $Y_C \equiv N\left(0, \int_0^{\infty} e^{-pC} \Omega_{xx} e^{-pC} dp\right)$ *as defined in Lemma A.9.*

The limit distribution of $\hat{B}^{LHIVX}$ with mildly explosive regressors is therefore essentially the same as the OLS estimator without using IVX. In short horizon models this result was confirmed in Phillips and Lee (2012b).

The variance matrix in (3.18) can be estimated leading to a self-normalized estimator that is robust to mildly explosive regressors. As the following result shows we therefore have a single uniform inference method that covers all of the cases (I1)–(I5).

Theorem 3.6.

$$vec\left\{\left(\hat{B}^{LHIVX} - B\right)\right\}'\left[\left(X'P_{\tilde{Z}}X\right)^{-1} \otimes \hat{\Omega}_{00}\right]^{-1} \times vec\left\{\left(\hat{B}^{LHIVX} - B\right)\right\} \Longrightarrow \chi^2(mp).$$

Theorem 3.6 shows that, although the LHIVX estimator with mildly explosive regressors has different asymptotic behavior than the other cases (I1)–(I4), the self-normalized estimator has the same standard chi square limit theory which is pivotal and simple to use. Theorems 3.2, 3.4 and 3.6 together show that a single test procedure for predictive regression may be employed that covers a large range of potential predictors with autoregressive roots in a wide neighborhood of unity.

4. Conclusion

This paper develops a theory of inference for long-horizon predictive regressions with multivariate regressors that have roots in a broad vicinity of unity. We propose a long-horizon version of the IVX method which extends some recent developments on this methodology in Magdalinos and Phillips (2009b). Our procedure allows the autoregressive roots of the predictive regressors to range from mildly integrated (close to the boundary of stationarity) through to mildly explosive roots (close to the boundary of full explosive behavior). This extension is more comprehensive than existing local to unity specifications for predictive regressors.

The method developed here has both empirical and theoretical benefits. The LHIVX test statistics have asymptotic chi-square distributions that are free of any nuisance parameters and there is no need for simulation methods. The method also has non-trivial power against local alternatives. These advantages are especially appealing for empirical practice in view of the simplicity, convenience and generality of the approach. A further notable feature is that the predictive regressions can be multivariate with regard to both regressor and regressand, thereby extending the range of coverage and realism in empirical model testing of predictability. To the best of our knowledge, this general formulation of predictive regression is not presently available in the literature.

Acknowledgment

The first author's research support is acknowledged from NSF grant no. SES 09-56687 and from the University of Auckland.

Appendix

This Section provides some useful preliminaries and proofs of the main results in the paper. An online supplement (Phillips and Lee, 2012a) contains proofs of the supporting lemmas listed here and further technical arguments that are used in the paper. Also included is a discussion of model validity and equation balancing in predictive regression under various alternatives.

A.1. Notation, conditions and limit theory of Section 2

We provide conditions for the results given in the review Section 2 and the accompanying limit theory. We impose the following moment conditions on the mds innovations in (2.1) and (2.2),

$$u_t := \begin{bmatrix} u_{0t} \\ u_{xt} \end{bmatrix}, \qquad \mathbb{E}_{\mathcal{F}_{t-1}}\left[u_t u_t'\right] = \Sigma = \begin{bmatrix} \Sigma_{00} & \Sigma_{0x} \\ \Sigma_{x0} & \Sigma_{xx} \end{bmatrix} > 0,$$

$$\sup_t \mathbb{E}\left[u_{it}^4\right] < \infty \quad \text{for } i = 0, x$$

where the natural filtration is $\mathcal{F}_t = \sigma(x_0, y_0, u_t, u_{t-1}, \ldots)$. Under this innovation framework, we have the well known functional law (e.g. Phillips and Solo, 1992)

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor ns \rfloor} u_j := \frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor ns \rfloor} \begin{bmatrix} u_{0j} \\ u_{xj} \end{bmatrix} = \begin{bmatrix} B_{0n}(s) \\ B_{xn}(s) \end{bmatrix} \Longrightarrow \begin{bmatrix} B_0(s) \\ B_x(s) \end{bmatrix} = BM \begin{bmatrix} \Sigma_{00} & \Sigma_{0x} \\ \Sigma_{x0} & \Sigma_{xx} \end{bmatrix},$$

and local to unity limit law (Phillips, 1987) $\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} \Longrightarrow J_x^c(r)$, where $J_x^c(r) = \int_0^r e^{(r-s)C} dB_x(s)$ is a standard Ornstein–Uhlenbeck process. Under certain conditions, similar results for weakly dependent vector processes also hold (see Section 3.1).

A.2. Model validity and consistency under local alternatives

A common characteristic of linear predictive regression equations is their unbalanced nature under the alternative hypothesis. Stock return regressions are a striking example because the dependent variable (stock returns) has features close to an mds whereas the posited predictive regressors are typically persistent and are often modeled as near integrated processes. In consequence, most of the existing literature studies test performance and asymptotics only under the null hypothesis, completely forgoing problems of imbalance. Given that the focus is actually the alternative — constituting a search for regressors with predictive capability — attention to the null is really just a concern about controlling size, while the main interest is fundamentally on the power to detect predictive capability in potential explanatory variables.

This problem of imbalance is addressed here. We discuss how linear regression specification may be justified and the underlying model validated (or balanced) by assigning localized forms to the regression coefficients A and B in (3.1) and (3.10). This formulation allows the dependent variable to faithfully retain a near mds or $I(0)$ form even under the alternative, thereby ensuring model validity. Localized coefficients also enable us to analyze test power and determine conditions under which non negligible power may be achieved.

To avoid lengthy enumeration of the various cases we restrict our attention in what follows to cases (I2)–(I4). Local departures from the null based on model (3.1) may be expressed as

$$y_{t+1} = Ax_t + u_{0t+1}, \qquad \text{(A.1)}$$
$$\mathcal{H}_1 : A = A_n = \frac{a}{n^\gamma} \quad \text{for some } a, \gamma > 0.$$

In view of (3.8), this formulation results in the following sequence of local-to-zero predictive components from the regressors. First, we have

$$\mathbb{E}_{t+k-1}\left[y_{t+k}\right] = \frac{a}{n^\gamma} x_{t+k-1},$$
$$\mathbb{E}_{t+k-2}\left[y_{t+k}\right] = \frac{a}{n^\gamma}\mathbb{E}_{t+k-2}\left[x_{t+k-1}\right] = \frac{a}{n^\gamma}\left[\left(1+\frac{c}{n}\right)x_{t+k-2} + u_{xt+k-1}\right],$$
$$\vdots$$
$$\mathbb{E}_t\left[y_{t+k}\right] = \frac{a}{n^\gamma}\left[\left(1+\frac{c}{n}\right)^{k-1} x_t + \sum_{j=1}^{k-1}\left(1+\frac{c}{n}\right)^{k-1-j} u_{xt+j}\right] = \frac{a}{n^\gamma}x_t + O_p\left(\frac{\sqrt{k}}{n^\gamma}\right). \qquad \text{(A.2)}$$

Summing up we have

$$\begin{aligned}\sum_{i=1}^k \mathbb{E}_{t+k-i}\left[y_{t+k}\right] &= \frac{a}{n^\gamma}\left[\sum_{i=1}^k\left(1+\frac{c}{n}\right)^{i-1} x_{t+k-j} + \sum_{i=1}^k\sum_{j=1}^{i-1}\left(1+\frac{c}{n}\right)^{i-1-j} u_{xt+j}\right] \\ &= \frac{a}{n^\gamma}\left[\left(\sum_{i=1}^k x_{t+k-j}\right)\left\{1+o_p(1)\right\} + \sum_{i=1}^k i^{1/2}\left(i^{-1/2}\sum_{j=1}^{i-1}\left(1+\frac{c}{n}\right)^{i-1-j} u_{xt+j}\right)\right] \\ &= \frac{a}{n^\gamma}\left[x_t^k\left\{1+o_p(1)\right\} + O_p\left(\sqrt{k}\right)\right] \\ &= \frac{a}{n^\gamma}x_t^k\left\{1+o_p(1)\right\} = k\mathbb{E}_t\left[y_{t+k}\right]\left\{1+o_p(1)\right\},\end{aligned}$$

using (A.2), which implies that

$$\mathbb{E}_t\left[y_{t+k}\right] = \frac{a}{kn^\gamma}x_t^k\left\{1+o_p(1)\right\}. \qquad \text{(A.3)}$$

These results enable us to investigate model validity (or balance) under the alternative hypothesis. Observe that under the alternative $\mathcal{H}_1$ (3.10) becomes

$$y_{t+k} = \frac{a}{kn^\gamma}x_t^k\left\{1+o_p(1)\right\} + u_{0t+k} = a\frac{kn^{1/2}}{kn^\gamma}\left(\frac{x_t^k}{kn^{1/2}}\right) + u_{0t+k} = u_{0t+k} + O_p\left(\frac{n^{\frac{1}{2}}}{n^\gamma}\right), \qquad \text{(A.4)}$$

Hence, to retain the validity of the regression equation (3.10) under the local departure, $\gamma > \frac{1}{2}$ preserves the apparent mds character of y_{t+k} and $\gamma \geq \frac{1}{2}$ preserves its $O_p(1)$ order of magnitude.

It might initially appear that local alternatives like $\mathcal{H}_1$ might be indistinguishable from the null $\mathcal{H}_0$. However, tests based on the long horizon IVX estimator $\hat{B}^{LHIVX}$ will have discriminatory power against local alternatives of the form $B = B_n = \frac{a}{kn^\gamma}$. In particular, we have

$$n^{\frac{1}{2}}k^{\frac{3}{2}}\left(\hat{B}^{LHIVX}\right) = n^{\frac{1}{2}}k^{\frac{3}{2}}\left(\hat{B}^{LHIVX} - B\right) + n^{\frac{1}{2}}k^{\frac{3}{2}}B = O_p(1) + O_p\left(\frac{k^{\frac{1}{2}}n^{\frac{1}{2}}}{n^\gamma}\right) \to \infty \quad \text{(A.5)}$$

provided $\gamma < \frac{1}{2} + \frac{\nu}{2}$ when $k = n^\nu$. Thus, the simple rate condition $\frac{1}{2} \le \gamma < \frac{1}{2} + \frac{\nu}{2}$ ensures (i) model validity and (ii) consistent testing against local alternatives using $\hat{B}^{LHIVX}$. In effect, local departures from the null must be small enough to preserve the character of y_{t+k} (including apparent mds behavior if that is evident), which is assured by the inequality $\frac{1}{2} \le \gamma$, and the forecast horizon must be large enough to ensure that the test statistic diverges under the alternative, which is assured by the inequality $\gamma < \frac{1}{2} + \frac{\nu}{2}$. Note that this formulation may in part explain some of the practical difficulty that has been encountered in establishing empirical evidence of predictability – the departures from the null are necessarily small in order to preserve the observed character of the dependent variable.

In a similar way, it can be demonstrated that the short-horizon estimator $\hat{B}^{IVX}$ (MP) can successfully discriminate the null from the local alternatives (which retain model validity) of the form $A_n = \frac{a}{n^\gamma}$ where $\frac{1}{2} \le \gamma < \frac{1}{2} + \frac{\delta}{2}$. From the rate condition $\delta < \nu$, we confirm that there is an asymptotic gain in discriminatory power by using the long-horizon IVX procedure. In effect, the use of a longer horizon (larger ν) enables higher discriminatory power thereby providing a central motivation for the use of long-horizon regression.

A.3. Useful lemmas

Hereafter, we use the spectral norm

$$\|M\| = \max_i \left\{\lambda_i^{1/2} : \lambda_i \text{ is an eigenvalue of } M'M\right\},$$

and other norms, such as the L_1 or L_2-norm, will be explicitly specified.

A.3.1. The $\alpha = 1$ case

The following lemmas develop the limit theory for the numerator and denominator elements in the LHIVX estimator (3.15) in Section 3.3.

Lemma A.1. 1. $\frac{C_z}{kn^{\frac{1}{2}+\delta}}\psi_{nt}^k \Longrightarrow J_x^c(r)$, *with* $t = \lfloor nr \rfloor$.
2. $\frac{1}{k^{\frac{1}{2}}n^{\frac{3}{2}+\delta}}\sum_{t=1}^{n-k} u_{0t+k}\left(\psi_{nt}^k\right)' C = o_p(1)$.
3. $\frac{C_z}{k^{\frac{1}{2}}n^\delta} z_t^k \Longrightarrow V_x(t) \equiv N(0, \Omega_{xx})$ *for any* t.
4. $\frac{1}{n}\sum_{t=1}^{n-k}\left(\frac{C_z}{k^{\frac{1}{2}}n^\delta}z_t^k\right)\left(\frac{C_z}{k^{\frac{1}{2}}n^\delta}z_t^k\right)' \to_p \Omega_{xx} = F_x(1)\Sigma F_x(1)'$.

From (3.17) the normalized numerator of $\hat{B}^{LHIVX}$ can be decomposed as

$$\frac{1}{k^{\frac{1}{2}}n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n-k} u_{0t+k}\left(\tilde{z}_t^k\right)' = \frac{1}{k^{\frac{1}{2}}n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n-k} u_{0t+k}\left(z_t^k\right)' + \frac{1}{k^{\frac{1}{2}}n^{\frac{3}{2}+\delta}}\sum_{t=1}^{n-k} u_{0t+k}\left(\psi_{nt}^k\right)' C.$$

Lemma A.1-2 shows the last term is negligible, thereby removing the dependence on C asymptotically. The limit theory of $k^{-\frac{1}{2}}n^{-(\frac{1}{2}+\delta)}\sum_{t=1}^{n-k} u_{0t+k}\left(z_t^k\right)'$ follows from Lemma A.1-3 and -4 and the martingale CLT.

Lemma A.2 (*Numerator of* $\hat{B}^{LHIVX}$).

$$Vec\left\{\frac{1}{k^{\frac{1}{2}}n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n-k} u_{0t+k}\left(\tilde{z}_t^k\right)'\right\} \Longrightarrow N\left(0, C_z^{-1}\Omega_{xx}C_z^{-1} \otimes \Omega_{00}\right).$$

Thus, the normalized numerator of $\hat{B}^{LHIVX}$ has a centered asymptotic normal distribution that does not depend on the nuisance parameter C but rather depends on the localizing coefficient C_z chosen in the construction of the IVX instruments. This removal of unestimable nuisance parameters in the limit distribution is one of the features of IVX and this benefit applies in long-horizon predictive regressions.

For the denominator matrix of $\hat{B}^{LHIVX}$ we have the following decomposition

$$\frac{1}{k^2n^{1+\delta}}\sum_{t=1}^{n-k} x_t^k\left(\tilde{z}_t^k\right)' = \frac{1}{k^2n^{1+\delta}}\sum_{t=1}^{n-k} x_t^k\left(z_t^k\right)' + \frac{1}{k^2n^{2+\delta}}\sum_{t=1}^{n-k} x_t^k\left(\psi_{nt}^k\right)' C',$$

and the following lemma gives the asymptotic behavior of these two normalized components and the overall matrix.

Lemma A.3. 1. $\frac{1}{k^2n^{2+\delta}}\sum_{t=1}^{n-k} x_t^k\left(\psi_{nt}^k\right)' C \Longrightarrow \left(\int_0^1 J_x^c(r)J_x^c(r)'dr\right) C_z^{-1}C$.
2. $\frac{1}{k^2n^{1+\delta}}\sum_{t=1}^{n-k} x_t^k\left(z_t^k\right)' C_z \to_p \frac{1}{2}\Omega_{xx}$.
3. (Denominator of $\hat{B}^{LHIVX}$) $\frac{1}{k^2n^{1+\delta}}\sum_{t=1}^{n-k} x_t^k\left(\tilde{z}_t^k\right)' \Longrightarrow \frac{1}{2}\Omega_{xx}C_z^{-1} + \left(\int_0^1 J_x^c(r)J_x^c(r)'dr\right)C_z^{-1}C$.

The next lemma helps establish an intermediate result for the self-normalized estimator in Theorem 3.2.

Lemma A.4. $\frac{1}{kn^{1+2\delta}}\sum_{t=1}^{n-k}\left(\tilde{z}_t^k\right)\left(\tilde{z}_t^k\right)' = \frac{1}{kn^{1+2\delta}}\sum_{t=1}^{n-k}\left(z_t^k\right)\left(z_t^k\right)' + o_p(1)$.

A.3.2. Mildly integrated regressors

We collect together the lemmas needed for Section 3.4. The normalized numerator in the LHIVX estimation error (3.15) is

$$\frac{1}{k^{\frac{1}{2}}n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n-k} u_{0t+k}\left(\tilde{z}_t^k\right)' = \frac{1}{k^{\frac{1}{2}}n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n-k} u_{0t+k}\left(z_t^k\right)' + \frac{1}{k^{\frac{1}{2}}n^{\frac{1}{2}+\alpha+\delta}}\sum_{t=1}^{n-k} u_{0t+k}\left(\psi_{nt}^k\right)' C,$$

with a dominant leading term and secondary term that is negligible. The following lemma establishes the negligibility of this second term.

Lemma A.5. $\frac{1}{k^{\frac{1}{2}}n^{\frac{1}{2}+\alpha+\delta}}\sum_{t=1}^{n-k} u_{0t+k}\left(\psi_{nt}^k\right)' = o_p(1)$.

Therefore, asymptotics of the normalized numerator are the same as the (I2)–(I4) cases and so are free of the nuisance parameter C.

Lemma A.6 (*Numerator of* $\hat{B}^{LHIVX}$).

$$vec\left\{\frac{1}{k^{\frac{1}{2}}n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n-k} u_{0t+k}\left(z_t^k\right)'\right\} \Longrightarrow N\left(0, C_z^{-1}\Omega_{xx}C_z^{-1} \otimes \Omega_{00}\right).$$

Now we turn to the denominator. It turns out we can remove the nuisance parameter dependence for the denominator as well. Hence, for the mildly integrated regressor case, we have a pivotal test statistic even without self-normalization.

Lemma A.7. 1. $\frac{1}{k^2 n^{1+\delta}} \sum_{t=1}^{n-k} x_t^k \left(z_t^k\right)' C_z \to_p \frac{1}{2}\Omega_{xx}$.
2. $\frac{1}{k^2 n^{1+\alpha+\delta}} \sum_{t=1}^{n-k} x_t^k \left(\psi_{nt}^k\right)' = o_p(1)$.

This result shows that a more concise limit theory holds for the denominator in the mildly integrated case.

$$\frac{1}{k^2 n^{1+\delta}} \sum_{t=1}^{n-k} x_t^k \left(\tilde{z}_t^k\right)' = \frac{1}{k^2 n^{1+\delta}} \sum_{t=1}^{n-k} x_t^k \left(z_t^k\right)' + \frac{1}{k^2 n^{1+\alpha+\delta}} \sum_{t=1}^{n-k} x_t^k \left(\psi_{nt}^k\right)' C' \to_p \frac{1}{2}\Omega_{xx} C_z^{-1}.$$

Lemma A.8 (*Denominator of* $\hat{B}^{LHIVX}$).

$$\frac{1}{k^2 n^{1+\delta}} \sum_{t=1}^{n-k} x_t^k \left(\tilde{z}_t^k\right)' \to_p \frac{1}{2}\Omega_{xx} C_z^{-1}.$$

A.3.3. Mildly explosive regressors

The lemmas used in Section 3.5 are developed here. The following is from Magdalinos and Phillips (2009a, Lemma 4.1).

Lemma A.9. *For each sequence* l_n *satisfying* $\|R_n\|^{-l_n} \to 0, n^\alpha \|R_n\|^{-(n-l_n)} \to 0$*, define* $Y_{Cn} := \frac{1}{n^{\alpha/2}} \sum_{j=1}^{l_n} R_n^{-j} F_x(1)\varepsilon_j$*. Then*

$$n^{-\alpha/2} R_n^{-n} x_n = \frac{1}{n^{\alpha/2}} \sum_{j=1}^{n} R_n^{-j} u_{xj} = \frac{1}{n^{\alpha/2}} \sum_{j=1}^{l_n} R_n^{-j} u_{xj} + o_p(1) = Y_{Cn} + o_p(1),$$

$$Y_{Cn} \Longrightarrow Y_C \equiv N\left(0, \int_0^\infty e^{-pC} \Omega_{xx} e^{-pC} dp\right).$$

The next lemma comes from other work by the authors (Phillips and Lee, 2012b, Lemma 5.4).

Lemma A.10. $\frac{1}{n^{\frac{\alpha}{2}+\delta}} R_n^{-t} \psi_{nt} = C_z^{-1} Y_C + o_p(1)$ *for all* $t \in [n^{\alpha'} + n^{\delta'}, n]$ *with* $\frac{n^\alpha}{n^{\alpha'}} + \frac{n^\delta}{n^{\delta'}} \to 0$.

As discussed in the text, the order of magnitude of ψ_{nt} becomes larger for explosive processes and ends up dominating the IVX asymptotics. Similarly, the LHIVX remainder ψ_{nt}^k also dominates the asymptotics. These characteristics are demonstrated in the following lemma.

Lemma A.11. 1. $\frac{1}{k n^{\frac{\alpha}{2}+\delta}} R_n^{-t} \psi_{nt}^k = C_z^{-1} Y_C + o_p(1)$ *for all* $t \in [n^{\alpha'} + n^{\delta'}, n-k]$.
2. $\frac{1}{k n^{\delta}} \sum_{t=1}^{n-k} u_{0t+k} \left(\tilde{z}_t^k\right)' R_n^{-n} = \frac{1}{k n^{\alpha+\delta}} \sum_{t=1}^{n-k} u_{0t+k} \left(\psi_{nt}^k\right)' R_n^{-n} C + o_p(1)$.
3. $vec\left(\frac{1}{k n^{\alpha+\delta}} \sum_{t=1}^{n-k} u_{0t+k} \left(\psi_{nt}^k\right)' R_n^{-n} C\right) \Longrightarrow \left(C_z^{-1} C \otimes I_m\right) \times MN\left(0, \int_0^\infty e^{-pC} Y_C Y_C' e^{-pC} dp \otimes \Omega_{00}\right)$.

It therefore follows that the asymptotic distribution of the normalized numerator directly comes from the limit behavior of the remainder term as shown in Lemma A.11-3.

Lemma A.12 (*Numerator of* $\hat{B}^{LHIVX}$).

$$vec\left(\frac{1}{k n^\delta} \sum_{t=1}^{n-k} u_{0t+k} \left(\tilde{z}_t^k\right)' R_n^{-n}\right) \Longrightarrow \left(C_z^{-1} C \otimes I_m\right) \times MN\left(0, \int_0^\infty e^{-pC} Y_C Y_C' e^{-pC} dp \otimes \Omega_{00}\right).$$

This result shows the nuisance parameter matrix C is not removed from the numerator if the regressors are mildly explosive. The coefficient $\frac{C}{n^\alpha}$ is not strong enough to make the remainder term negligible. Nonetheless the stronger signal of the mildly explosive regressors does enable the mixed normal limit theory of the LHIVX estimator. Eventually, therefore, we end up with a pivotal test statistic using self-normalization, as explained in the text.

The next lemma gives the asymptotic behavior of the components of the denominator.

Lemma A.13. 1. $\frac{1}{k n^{\alpha/2}} R_n^{-t} x_t^k = Y_C + o_p(1)$ *for all* $t \in [n^{\alpha'} + n^{\delta'}, n-k]$.
2. $\frac{1}{k^2 n^{\alpha+\delta}} \sum_{t=1}^{n-k} R_n^{-n} x_t^k \left(\tilde{z}_t^k\right)' R_n^{-n} = \frac{1}{k^2 n^{2\alpha+\delta}} \sum_{t=1}^{n-k} R_n^{-n} x_t^k \left(\psi_{nt}^k\right)' C' R_n^{-n} + o_p(1)$.
3. $\frac{1}{k^2 n^{2\alpha+\delta}} \sum_{t=1}^{n-k} R_n^{-n} x_t^k \left(\psi_{nt}^k\right)' R_n^{-n} \Longrightarrow \int_0^\infty e^{-pC} Y_C Y_C' e^{-pC} dp C_z^{-1}$

From the decomposition $\sum_{t=1}^{n-k} x_t^k \left(\tilde{z}_t^k\right)' = \sum_{t=1}^{n-k} x_t^k \left(z_t^k\right)' + n^{-\alpha} \sum_{t=1}^{n-k} x_t^k \left(\psi_{nt}^k\right)' C'$, we have

$$\frac{1}{k^2 n^{\alpha+\delta}} \sum_{t=1}^{n-k} R_n^{-n} x_t^k \left(\tilde{z}_t^k\right)' R_n^{-n} = \frac{1}{k^2 n^{2\alpha+\delta}} \sum_{t=1}^{n-k} R_n^{-n} x_t^k \left(\psi_{nt}^k\right)' R_n^{-n} C + o_p(1) \Longrightarrow \int_0^\infty e^{-pC} Y_C Y_C' e^{-pC} dp C_z^{-1} C.$$

As in the numerator case, the component involving the remainder plays the leading role in the asymptotics. Hence the limit of the normalized denominator involves the nuisance parameter C as well.

Lemma A.14 (*Denominator of* $\hat{B}^{LHIVX}$).

$$\frac{1}{k^2 n^{\alpha+\delta}} \sum_{t=1}^{n-k} R_n^{-n} x_t^k \left(\tilde{z}_t^k\right)' R_n^{-n} \Longrightarrow \int_0^\infty e^{-pC} Y_C Y_C' e^{-pC} dp C_z^{-1} C.$$

The following lemma is used in developing the limit theory for the self-normalized estimator given in Theorem 3.6.

Lemma A.15. $\frac{1}{k^2 n^{2\delta}} \sum_{t=1}^{n-k} R_n^{-n} \left(\tilde{z}_t^k\right) \left(\tilde{z}_t^k\right)' R_n^{-n} \Longrightarrow C C_z^{-1} \int_0^\infty e^{-pC} Y_C Y_C' e^{-pC} dp C_z^{-1} C$.

A.4. Proofs of the main results

Proof of Theorem 3.1. From Lemmas A.2 and A.3-3,

$$vec\left\{\frac{1}{k^{\frac{1}{2}} n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n-k} u_{0t+k} \left(z_t^k\right)'\right\} \Longrightarrow N\left(0, C_z^{-1} \Omega_{xx} C_z^{-1} \otimes \Omega_{00}\right),$$

$$\frac{1}{k^2 n^{1+\delta}} \sum_{t=1}^{n-k} x_t^k \left(\tilde{z}_t^k\right)' \Longrightarrow \frac{1}{2}\Omega_{xx} C_z^{-1} + \left(\int_0^1 J_c^x(r) J_c^x(r)' dr\right) C_z^{-1} C := \Psi_{cxz}.$$

Hence

$$vec\left\{n^{\frac{1}{2}}k^{\frac{3}{2}}\left(\hat{B}^{LHIVX}-B\right)\right\}$$
$$=\left[\left\{\left(\frac{1}{k^2n^{1+\delta}}\sum_{t=1}^{n-k}x_t^k\left(\tilde{z}_t^k\right)'\right)^{-1}\right\}'\otimes I\right]$$
$$\times Vec\left\{\frac{1}{k^{\frac{1}{2}}n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n-k}u_{0t+k}\left(z_t^k\right)'\right\}$$
$$\Longrightarrow MN\left(0,\left(\Psi_{cxz}^{-1}\right)'C_z^{-1}\Omega_{xx}C_z^{-1}\left(\Psi_{cxz}^{-1}\right)\otimes\Omega_{00}\right).$$

Mixed normality holds since the MG components of the numerator and denominator are asymptotically independent. In particular, defining the martingale difference

$$\xi_{nt+k}:=\begin{bmatrix}\frac{1}{k^{\frac{1}{2}}n^{\frac{1}{2}+\delta}}z_t^k\otimes\varepsilon_{t+k}\\ \frac{1}{\sqrt{n}}\varepsilon_{t+k}\end{bmatrix},$$

and letting $\mathcal{F}_{nt}$ be the natural filtration associated with the array ξ_{nt}, we have the martingale conditional variance

$$\sum\mathbb{E}_{\mathcal{F}_{nt+k-1}}\xi_{nt+k}\xi_{nt+k}'$$
$$=\begin{bmatrix}\frac{1}{n}\sum_{t=1}^{n-k}\left(\frac{1}{k^{\frac{1}{2}}n^{\delta}}z_t^k\right)\left(\frac{1}{k^{\frac{1}{2}}n^{\delta}}z_t^k\right)'\otimes\Sigma & \frac{1}{n}\sum_{t=1}^{n-k}\left(\frac{1}{k^{\frac{1}{2}}n^{\delta}}z_t^k\right)\otimes\Sigma\\ \frac{1}{n}\sum_{t=1}^{n-k}\left(\frac{1}{k^{\frac{1}{2}}n^{\delta}}z_t^k\right)\otimes\Sigma & \Sigma\end{bmatrix}$$
$$\to_p\begin{bmatrix}C_z^{-1}\Omega_{xx}C_z^{-1}\otimes\Sigma & 0\\ 0 & \Sigma\end{bmatrix},$$

from Lemma A.1-3 and -4. □

Proof of Theorem 3.2. From Lemmas A.1-4 and A.4, we know that

$$\frac{1}{n}\sum_{t=1}^{n-k}\left(\frac{C_z}{k^{\frac{1}{2}}n^{\delta}}\tilde{z}_t^k\right)\left(\frac{C_z}{k^{\frac{1}{2}}n^{\delta}}\tilde{z}_t^k\right)'$$
$$=\frac{1}{n}\sum_{t=1}^{n-k}\left(\frac{C_z}{k^{\frac{1}{2}}n^{\delta}}z_t^k\right)\left(\frac{C_z}{k^{\frac{1}{2}}n^{\delta}}z_t^k\right)'+o_p(1)=\Omega_{xx}+o_p(1).$$

Moreover, $\frac{1}{k^2n^{1+\delta}}\sum_{t=1}^{n-k}x_t^k\left(\tilde{z}_t^k\right)'\Longrightarrow\Psi_{cxz}$. Thus,

$$nk^3\left(X'P_{\tilde{Z}}X\right)^{-1}=\left\{\left(\frac{1}{k^2n^{1+\delta}}\sum_{t=1}^{n-k}x_t^k\left(\tilde{z}_t^k\right)'\right)\right.$$
$$\left.\times\left(\frac{1}{kn^{1+2\delta}}\sum_{t=1}^{n-k}\left(\tilde{z}_t^k\right)\left(\tilde{z}_t^k\right)'\right)^{-1}\left(\frac{1}{k^2n^{1+\delta}}\sum_{t=1}^{n-k}x_t^k\left(\tilde{z}_t^k\right)'\right)'\right\}^{-1}$$
$$\Longrightarrow\left\{\left(\Psi_{cxz}\right)C_z\left(\Omega_{xx}\right)^{-1}C_z\left(\Psi_{cxz}\right)'\right\}^{-1}$$
$$=\left(\Psi_{cxz}^{-1}\right)'C_z^{-1}\left(\Omega_{xx}\right)C_z^{-1}\left(\Psi_{cxz}^{-1}\right),$$

and

$$\left[nk^3\left(X'P_{\tilde{Z}}X\right)^{-1}\otimes\hat{\Omega}_{00}\right]^{-1}$$
$$\Longrightarrow\left[\left(\Psi_{cxz}^{-1}\right)'C_z^{-1}\left(\Omega_{xx}\right)C_z^{-1}\left(\Psi_{cxz}^{-1}\right)\otimes\Omega_{00}\right]^{-1}=\Sigma_B^{-1},$$

giving the required result. □

Proof of Theorem 3.3. From Lemmas A.6 and A.8, we have

$$vec\left\{n^{\frac{1}{2}}k^{\frac{3}{2}}\left(\hat{B}^{LHIVX}-B\right)\right\}$$
$$=\left[\left\{\left(\frac{1}{k^2n^{1+\delta}}\sum_{t=1}^{n-k}x_t^k\left(\tilde{z}_t^k\right)'\right)^{-1}\right\}'\otimes I\right]$$
$$\times vec\left\{\frac{1}{k^{\frac{1}{2}}n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n-k}u_{0t+k}\left(z_t^k\right)'\right\}$$
$$\Longrightarrow\left[\left\{\left(\frac{1}{2}\Omega_{xx}C_z^{-1}\right)^{-1}\right\}'\otimes I\right]N\left(0,C_z^{-1}\Omega_{xx}C_z^{-1}\otimes\Omega_{00}\right)$$
$$\equiv N\left(0,4\Omega_{xx}^{-1}\otimes\Omega_{00}\right).\quad\square$$

Proof of Theorem 3.4. From Lemmas A.1-4 and A.4, we have

$$\frac{1}{n}C_z^{-1}\sum_{t=1}^{n-k}\left(\frac{C_z}{k^{\frac{1}{2}}n^{\delta}}\tilde{z}_t^k\right)\left(\frac{C_z}{k^{\frac{1}{2}}n^{\delta}}\tilde{z}_t^k\right)'C_z^{-1}$$
$$=\frac{1}{n}C_z^{-1}\sum_{t=1}^{n-k}\left(\frac{C_z}{k^{\frac{1}{2}}n^{\delta}}z_t^k\right)\left(\frac{C_z}{k^{\frac{1}{2}}n^{\delta}}z_t^k\right)'C_z^{-1}+o_p(1),$$
$$\to_p C_z^{-1}\Omega_{xx}C_z^{-1},$$

and from Lemma A.8,

$$\frac{1}{k^2n^{1+\delta}}\sum_{t=1}^{n-k}x_t^k\left(\tilde{z}_t^k\right)'\to_p\frac{1}{2}\Omega_{xx}C_z^{-1}.$$

Thus,

$$nk^3\left(X'P_{\tilde{Z}}X\right)^{-1}=\left\{\left(\frac{1}{k^2n^{1+\delta}}\sum_{t=1}^{n-k}x_t^k\left(\tilde{z}_t^k\right)'\right)\right.$$
$$\left.\times\left(\frac{1}{kn^{1+2\delta}}\sum_{t=1}^{n-k}\left(\tilde{z}_t^k\right)\left(\tilde{z}_t^k\right)'\right)^{-1}\left(\frac{1}{k^2n^{1+\delta}}\sum_{t=1}^{n-k}x_t^k\left(\tilde{z}_t^k\right)'\right)'\right\}^{-1}$$
$$\to_p\left\{\frac{1}{2}\Omega_{xx}C_z^{-1}\left(C_z^{-1}\Omega_{xx}C_z^{-1}\right)^{-1}\left(\frac{1}{2}\Omega_{xx}C_z^{-1}\right)'\right\}^{-1}$$
$$=4\Omega_{xx}^{-1},$$

and then

$$\left[nk^3\left(X'P_{\tilde{Z}}X\right)^{-1}\otimes\hat{\Omega}_{00}\right]^{-1}\to_p\left[4\Omega_{xx}^{-1}\otimes\Omega_{00}\right]^{-1},$$

which leads to the required result. □

Proof of Theorem 3.5. From Lemmas A.12 and A.14,

$$vec\left(\frac{1}{kn^{\delta}}\sum_{t=1}^{n-k}u_{0t+k}\left(\tilde{z}_t^k\right)'R_n^{-n}\right)$$
$$\Longrightarrow\left(CC_z^{-1}\otimes I_m\right)\times MN\left(0,\int_0^\infty e^{-pC}Y_CY_C'e^{-pC}dp\otimes\Omega_{00}\right),$$
$$\frac{1}{k^2n^{\alpha+\delta}}\sum_{t=1}^{n-k}R_n^{-n}x_t^k\left(\tilde{z}_t^k\right)'R_n^{-n}\Longrightarrow\int_0^\infty e^{-pC}Y_CY_C'e^{-pC}dpC_z^{-1}C.$$

Joint convergence of these two processes is established in the context of OLS estimation by Magdalinos and Phillips (2009a, Proof of Theorem 4.1).

Hence

$$vec\left[kn^{\alpha}\left(\hat{B}^{LHIVX}-B\right)R_n^n\right]=vec\left[\left(\frac{1}{kn^{\delta}}\sum_{t=1}^{n-k}u_{0t+k}\left(\tilde{z}_t^k\right)'R_n^{-n}\right)\right.$$
$$\left.\times\left(\frac{1}{k^2n^{\alpha+\delta}}\sum_{t=1}^{n-k}R_n^{-n}x_t^k\left(\tilde{z}_t^k\right)'R_n^{-n}\right)^{-1}\right]$$
$$=\left[\left\{\left(\frac{1}{k^2n^{\alpha+\delta}}\sum_{t=1}^{n-k}R_n^{-n}x_t^k\left(\tilde{z}_t^k\right)'R_n^{-n}\right)'\right\}^{-1}\otimes I_m\right]$$
$$\times vec\left(\frac{1}{kn^{\delta}}\sum_{t=1}^{n-k}u_{0t+k}\left(\tilde{z}_t^k\right)'R_n^{-n}\right)$$
$$\Longrightarrow\left[\left\{\left(\int_0^{\infty}e^{-pC}Y_CY_C'e^{-pC}dpC_z^{-1}C\right)'\right\}^{-1}\otimes I_m\right]$$
$$\times\left(CC_z^{-1}\otimes I_m\right)\times MN\left(0,\int_0^{\infty}e^{-pC}Y_CY_C'e^{-pC}dp\otimes\Omega_{00}\right)$$
$$=\left[\left(\int_0^{\infty}e^{-pC}Y_CY_C'e^{-pC}dp\right)^{-1}\otimes I_m\right]$$
$$\times MN\left(0,\int_0^{\infty}e^{-pC}Y_CY_C'e^{-pC}dp\otimes\Omega_{00}\right)$$
$$\equiv MN\left(0,\left(\int_0^{\infty}e^{-pC}Y_CY_C'e^{-pC}dp\right)^{-1}\otimes\Omega_{00}\right),$$

as required. □

Proof of Theorem 3.6. From Lemmas A.14 and A.15, we know

$$\frac{1}{k^2n^{\alpha+\delta}}\sum_{t=1}^{n-k}R_n^{-n}x_t^k\left(\tilde{z}_t^k\right)'R_n^{-n}\Longrightarrow\int_0^{\infty}e^{-pC}Y_CY_C'e^{-pC}dpC_z^{-1}C,$$
$$\frac{1}{k^2n^{2\delta}}\sum_{t=1}^{n-k}R_n^{-n}\left(\tilde{z}_t^k\right)\left(\tilde{z}_t^k\right)'R_n^{-n}\Longrightarrow CC_z^{-1}\int_0^{\infty}e^{-pC}Y_CY_C'e^{-pC}dpC_z^{-1}C.$$

Thus,

$$k^2n^{2\alpha}R_n^n\left(X'P_{\tilde{Z}}X\right)^{-1}R_n^n=\left\{\left(\frac{1}{k^2n^{\alpha+\delta}}\sum_{t=1}^{n-k}R_n^{-n}x_t^k\left(\tilde{z}_t^k\right)'R_n^{-n}\right)\right.$$
$$\times\left(\frac{1}{k^2n^{2\delta}}\sum_{t=1}^{n-k}R_n^{-n}\left(\tilde{z}_t^k\right)\left(\tilde{z}_t^k\right)'R_n^{-n}\right)^{-1}$$
$$\left.\times\left(\frac{1}{k^2n^{\alpha+\delta}}\sum_{t=1}^{n-k}R_n^{-n}x_t^k\left(\tilde{z}_t^k\right)'R_n^{-n}\right)'\right\}^{-1}$$
$$\Longrightarrow\left\{\left(\int_0^{\infty}e^{-pC}Y_CY_C'e^{-pC}dpC_z^{-1}C\right)\right.$$
$$\times\left(CC_z^{-1}\int_0^{\infty}e^{-pC}Y_CY_C'e^{-pC}dpC_z^{-1}C\right)^{-1}$$
$$\left.\times\left(\int_0^{\infty}e^{-pC}Y_CY_C'e^{-pC}dpC_z^{-1}C\right)'\right\}^{-1}$$
$$=\left(\int_0^{\infty}e^{-pC}Y_CY_C'e^{-pC}dp\right)^{-1}$$

and

$$\left[k^2n^{2\alpha}R_n^n\left(X'P_{\tilde{Z}}X\right)^{-1}R_n^n\otimes\hat{\Omega}_{00}\right]^{-1}$$
$$\Longrightarrow\left[\left(\int_0^{\infty}e^{-pC}Y_CY_C'e^{-pC}dp\right)^{-1}\otimes\Omega_{00}\right]^{-1}.$$

Therefore,

$$vec\left\{\left(\hat{B}^{LHIVX}-B\right)\right\}'\left[\left(X'P_{\tilde{Z}}X\right)^{-1}\otimes\hat{\Omega}_{00}\right]^{-1}vec\left\{\left(\hat{B}^{LHIVX}-B\right)\right\}$$
$$=vec\left\{\left(\hat{B}^{LHIVX}-B\right)\right\}'\left(R_n^n\otimes I_m\right)\left(R_n^n\otimes I_m\right)^{-1}$$
$$\times\left[\left(X'P_{\tilde{Z}}X\right)^{-1}\otimes\hat{\Omega}_{00}\right]^{-1}$$
$$\times\left(R_n^n\otimes I_m\right)^{-1}\left(R_n^n\otimes I_m\right)vec\left\{\left(\hat{B}^{LHIVX}-B\right)\right\}$$
$$=vec\left[kn^{\alpha}\left(\hat{B}^{LHIVX}-B\right)R_n^n\right]'$$
$$\times\left[k^2n^{2\alpha}R_n^n\left(X'P_{\tilde{Z}}X\right)^{-1}R_n^n\otimes\hat{\Omega}_{00}\right]^{-1}$$
$$vec\left[kn^{\alpha}\left(\hat{B}^{LHIVX}-B\right)R_n^n\right]$$
$$\Longrightarrow\chi^2\left(mp\right).\quad\square$$

References

Amihud, Y., Hurvich, C., 2004. Predictive regressions: a reduced-bias estimation method. Journal of Financial and Quantitative Analysis 39 (04), 813–841.

Campbell, J., Yogo, M., 2006. Efficient tests of stock return predictability. Journal of Financial Economics 81 (1), 27–60.

Cavanagh, C., Elliott, G., Stock, J., 1995. Inference in models with nearly integrated regressors. Econometric Theory 11 (05), 1131–1147.

Chan, N.H., Wei, C.Z., 1987. Asymptotic inference for nearly nonstationary AR(1) processes. Annals of Statistics 15, 1050–1063.

Cochrane, J., 1991. Volatility tests and efficient markets. Journal of Monetary Economics 27, 463–485.

Elliott, G., 2011. A control function approach for testing the usefulness of trending variables in forecast models and linear regression. Journal of Econometrics 164, 79–91.

Elliott, G., Rothenberg, T.J., Stock, J.H., 1996. Efficient tests for an autoregressive unit root. Econometrica 64, 813–836.

Elliott, G., Stock, J.H., 1994. Inference in time series regression when the order of integration of a regressor is unknown. Econometric Theory 10 (3–4), 672–700.

Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33 (1), 3–56.

Gonzalo, J., Pitarakis, J., 2009. Regime specific predictability in predictive regressions. Discussion Papers in Economics and Econometrics (0916), University of Southampton.

Gospodinov, N., 2009. A new look at the forward premium puzzle. Journal of Financial Econometrics 7 (03), 312–338.

Hjalmarsson, E., 2011. New methods for inference in long-horizon regressions. Journal of Financial and Quantitative Analysis 46 (3), 815–839.

Jansson, M., Moreira, M., 2006. Optimal inference in regression models with nearly integrated regressors. Econometrica 74 (3), 681–714.

Jegadeesh, N., 1991. Seasonality in stock price mean reversion: evidence from the US and the UK. Journal of Finance 1427–1444.

Kasparis, I., Andreou, E., Phillips, P.C.B., 2012. Nonparametric Predictive Regression. Unpublished manuscript, Yale University.

Kendall, M.G., 1954. Note on bias in the estimation of autocorrelation. Biometrika 41, 403–404.

Kostakis, A., Magdalinos, A., Stamatogiannis, M., 2010. Robust econometric inference for stock return predictability. Unpublished Manuscript, University of Nottingham.

Kothari, S.P., Shanken, J., 1997. Book-to-market, dividend yield, and expected market returns: a time series analysis. Journal of Financial Economics 18, 169–203.

Lee, J.H., 2012. Predictive quantile regression with persistent covariates. Unpublished Manuscript, Yale University.

Lewellen, J., 2004. Predicting returns with financial ratios. Journal of Financial Economics 74 (2), 209–235.

Liu, W., Maynard, A., 2007. A new application of exact nonparametric methods to long-horizon predictability tests. Studies in Nonlinear Dynamics and Econometrics 11 (1).

Magdalinos, T., Phillips, P.C.B., 2009a. Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. Econometric Theory 25, 482–526.
Magdalinos, T., Phillips, P.C.B., 2009b. Econometric inference in the vicinity of unity. CoFie Working Paper (7), Singapore Management University.
Pesaran, M.H., Timmermann, A., 1995. Predictability of stock returns: robustness and economic significance. Journal of Finance 50, 1201–1228.
Phillips, P.C.B., 1987. Towards a unified asymptotic theory for autoregression. Biometrika 74, 535–547.
Phillips, P.C.B., 2012a. Estimation of the localizing rate for mildly integrated and mildly explosive processes. Yale University mimeographed.
Phillips, P.C.B., 2012b. On confidence intervals for autoregressive roots and predictive regression. Unpublished Manuscript, CFDP 1879.
Phillips, P.C.B., Hansen, B., 1990. Statistical inference in instrumental variables regression with I(1) processes. The Review of Economic Studies 57 (1), 99.
Phillips, P.C.B., Lee, J.H., 2012a. Technical Supplement to: Predictive Regression under Various Degrees of Persistence and Robust Long-Horizon Regression. Online Technical Supplement. Available at https://sites.google.com/site/jihyung412/research.
Phillips, P.C.B., Lee, J.H., 2012b. Robust econometric inference with mixed integrated and mildly explosive regressors. Unpublished Manuscript, Yale University.
Phillips, P.C.B., Magdalinos, T., 2007a. Limit theory for moderate deviations from a unit root. Journal of Econometrics 136, 115–130.
Phillips, P.C.B., Magdalinos, T., 2007b. Limit theory for moderate deviations from a unit root under weak dependence. In: Phillips, G.D.A., Tzavalis, E. (Eds.), The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis. Cambridge University Press, Cambridge, pp. 123–162.
Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. The Annals of Statistics 971–1001.
Phillips, P.C.B., Wu, Y., Yu, J., 2011. Explosive behavior in the 1990s Nasdaq: when did exuberance escalate asset values? International Economic Review 52, 201–226.
Phillips, P.C.B., Yu, J., 2011. Dating the timeline of financial bubbles during the subprime crisis. Quantitative Economics 2, 455–491.
Stambaugh, R., 1999. Predictive regressions. Journal of Financial Economics 54 (375), 421.
Stock, J., 1991. Confidence intervals for the largest autoregressive root in US macroeconomic time series. Journal of Monetary Economics 28 (3), 435–459.
Torous, W., Valkanov, R., 2000. Boundaries of predictability: noisy predictive regressions. Unpublished Manuscript, UCLA.
Valkanov, R., 2003. Long-horizon regressions: theoretical results and applications. Journal of Financial Economics 68 (2), 201–232.